Gabor Fichtinger
Anne Martel
Terry Peters (Eds.)

LNCS 6891

Medical Image Computing and Computer-Assisted Intervention – MICCAI 2011

14th International Conference
Toronto, Canada, September 2011
Proceedings, Part I

MICCAI

Springer

Lecture Notes in Computer Science 6891

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

Gabor Fichtinger Anne Martel
Terry Peters (Eds.)

Medical Image Computing and Computer-Assisted Intervention – MICCAI 2011

14th International Conference
Toronto, Canada, September 18-22, 2011
Proceedings, Part I

Springer

Volume Editors

Gabor Fichtinger
Queen's University
Kingston, ON K7L 3N6, Canada
E-mail: gabor@cs.queensu.ca

Anne Martel
Sunnybrook Research Institute
Toronto, ON M4N 3M5, Canada
E-mail: anne.martel@sri.utoronto.ca

Terry Peters
Robarts Research Institute
London, ON N6A 5K8, Canada
E-mail: tpeters@robarts.ca

ISSN 0302-9743 e-ISSN 1611-3349
ISBN 978-3-642-23622-8 e-ISBN 978-3-642-23623-5
DOI 10.1007/978-3-642-23623-5
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2011935219

CR Subject Classification (1998): I.4, I.5, I.3.5-8, I.2.9-10, J.3, I.6

LNCS Sublibrary: SL 6 – Image Processing, Computer Vision, Pattern Recognition, and Graphics

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

The 14th International Conference on Medical Image Computing and Computer Assisted Intervention, MICCAI 2011, was held in Toronto, Canada during September, 18–22, 2011. The venue was the Westin Harbour Castle Hotel and Conference Centre on the waterfront of Lake Ontario in Downtown Toronto, the world's most ethnically diverse city.

MICCAI is the foremost international scientific event in the field of medical image computing and computer-assisted intervention. The annual conference has a high scientific standard by virtue of the threshold for acceptance, and accordingly MICCAI has built up a track record of attracting leading scientists, engineers and clinicians from a wide range of technical and biomedical disciplines. The year 2011 saw a record 819 paper submissions.

The Program Committee (PC) of MICCAI 2011 comprised 53 members. Each of the 819 papers was assigned to two PC members (a primary and a secondary) according to their expertise and the subject matter of the paper. The primary member knew the identity of the authors, but the secondary one did not. Each PC member had about 17 papers as primary and a further 17 as secondary member. The primary PC member assigned at least three external reviewers to each paper, according to their expertise and the subject matter of the paper. The external reviewers provided double-blind reviews of the papers, and authors were given the opportunity to rebut the anonymous reviews. In cases where reviewer opinions differed significantly and/or the rebuttal made it necessary, the primary member initiated a discussion among the reviewers. The primary member summarized the outcome of the discussion in a short report for the secondary. Finally, the secondary member considered all input (the reviews, rebuttal, discussion, primary's report, and, almost importantly, the paper itself) and made a recommendation for acceptance or rejection. The secondary PC member did not know the identity of the authors.

A two-day PC meeting was held with 33 of the PC members present. Each paper received fair consideration in a three-phase decision process.

- First stage: Initial acceptance of papers ranked very high by both the reviewers and the secondary PC member. Initial rejection of papers ranked very low by both the reviewers and the secondary PC member.
- Second stage: groups of five to seven PC members ranked the remaining papers and again selected the best papers and rejected the lowest ranking papers.
- Third stage: a different set of groups selected the best papers from the remaining undecided papers and rejected the rest.

The PC finally accepted 251 papers, giving a 30% acceptance rate.

We are greatly indebted to the reviewers and to the members of the PC for their extraordinary efforts assessing and evaluating the submissions within a very short time frame.

In 2011, attendees saw two changes in the way the program was organized. All accepted papers were presented as posters, and a subset of these were also invited for oral presentation, which were organized in clinical themes rather than by methodology as in earlier years. Poster sessions were organized in their traditional technical themes as in the past.

In addition to the main 3-day conference, the annual MICCAI event hosted an increased number of satellite tutorials and workshops, organized on the day before and the day after the main conference. This year's call for submission for tutorials and workshops led to a record 21 workshops and 8 tutorials accepted by a committee headed by Randy Ellis (Queen's University) and Purang Abolmaesumi (University of British Columbia). The tutorials provided a comprehensive overview of many areas in both the MIC and CAI domains, offering a unique educational forum for graduate students and postdoctoral fellows. The workshops presented an opportunity to present research, often in an early stage, to peer groups in a relaxed environment that allowed valuable discussion and feedback. The workshop subjects highlighted topics that were not all fully covered in the main conference, and thus added to the diversity of the MICCAI program.

In reviewing the proposals for these events, emphasis was given to workshop submissions that provided a comprehensive and interactive forum to address an open problem in MICCAI. We also promoted tutorials that related to an existing sub-discipline of MICCAI with known materials, approaches and open problems to help train new professionals in the field. Among the accepted workshops, several focused on emerging trends in the field of multi-modal statistical atlases, advanced computational and biomechanical models, and high-performance computing. MICCAI 2011 also hosted eight tutorials that spanned a wide spectrum of topics in basic and advanced software development for medical image analysis, algorithms for image segmentation, registration and visualization, as well as those highlighting new techniques in image-guided interventions. We would like to thank the Workshop and Tutorial Committee for their hard work in putting together such a comprehensive and unique program.

Two of the highlights of the conference were the keynote lectures by two Canadian scientists. Dafydd (Dave) Williams, physician, astronaut, medical robotics researcher, and recently, Hospital CEO, opened the conference with a presentation that looked at lessons that the health care system and medical researchers could learn from the challenges of space travel. The second keynote was given by Mark Henkleman, Director of the Mouse Imaging Centre, Toronto Centre for Phenogenomics, who spoke about high-throughput small-animal imaging techniques and quantitative statistical analysis methods for mapping phenotypic changes associated with genetic disease models in mice.

MICCAI 2011 would not have been feasible without the efforts of many people behind the scenes. We are particularly indebted to the local Organizing Committee in London and Toronto consisting of Janette Wallace, Johanne Guillemette,

Jackie Williams, Jade Orkin-Fenster, Debbie Lilley, Shuo Li, Perry Radau, and Raphael Ronen. In addition, we are deeply grateful to the Robarts Research Institute, the University of Western Ontario, Sunnybrook Research Institute, and Queen's University for their support in ensuring the success of this meeting, and to the staff at Springer for their continued high standards aimed at maintaining the MICCAI proceedings as the flagship of the LNCS series.

We thank the MICCAI Society Board for trusting us with the mandate to organize this conference, and to the Board and staff members for valuable and continuous advice and support through all phases of the project.

A special word of thanks goes to our sponsors, who generously provided financial support for the conference as a whole as well as for specific activities. This greatly assisted with the overall organization of the meeting, enabled us to continue offering best paper awards in various categories, and provided travel stipends to a significant number of student participants.

It was our great pleasure to welcome the attendees to Toronto for this year's MICCAI conference along with its satellite tutorials and workshops. Next year, the 15^{th} International Conference on Medical Image Computing and Computer-Assisted Intervention will be held in Nice, France, October 1–5, 2012. We look forward to seeing you all there.

September 2011

Gabor Fichtinger
Anne Martel
Terry Peters

Organization

Co-chairs

Gabor Fichtinger	Queen's University, Kingston, Canada
Anne Martel	Sunnybrook Research Institute, Toronto, Canada
Terry Peters	Robarts Research Institute, London, Canada

Workshop/Tutorial Co-chairs

Purang Abolmaesumi	University of British Columbia, Vancouver, Canada
Randy Ellis	Queen's University, Kingston, Canada

MICCAI Society, Board of Directors

James Duncan (President)	Yale University, Connecticut, USA
Gabor Fichtinger (Treasurer)	Queen's University, Canada
Alison Noble (Executive Director)	Oxford University, UK
Sebastien Ourselin (Secretary)	University College London, UK
Kensaku Mori	Nagoya University, Japan
Daniel Rueckert	Imperial College London, UK
Nicholas Ayache	INRIA, Sophia-Antipolis, France
Kevin Cleary	National Children's Hospital, Washington DC, USA
Polina Golland	MIT, Cambridge, USA
David Hawkes	University College London, UK
Tianzi Jiang	Chinese Academy of Sciences, Beijing, China
Wiro Niessen	Erasmus MC, Rotterdam, The Netherlands
Ichiro Sakuma	University of Tokyo, Japan
William (Sandy) Wells	Harvard Medical School, Boston, 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
Elsa Angelini	Telecom Paris, France
Stephen Aylward	Kitware, USA
Christian Barillot	INRIA, France
Wolfgang Birkfellner	Medical University of Vienna, Austria
Albert C.S. Chung	Hong Kong University of Science and Technology, China
Ela Claridge	University of Birmingham, UK
D. Louis Collins	McGill University, Canada
Dorin Comaniciu	Siemens Medical, USA
Christos Davatzikos	University of Pennsylvania, USA
Marleen de Bruijne	Erasmus Medical College, The Netherlands
Hervé Delingette	INRIA, France
James S. Duncan	Yale University, USA
Philip (Eddie) Edwards	Imperial College London, UK
Bernd Fischer	University of Lübeck, Germany
P. Thomas Fletcher	University of Utah, USA
Alejandro Frangi	Pompeu Fabra University, Spain
Polina Golland	Massachusetts Institute of Technology, USA
Nobuhiko Hata	Harvard Medical School, USA
David Hawkes	University College London, UK
David Holmes	Mayo Clinic, USA
Ameet Jain	Philips Medical Systems, USA
Pierre Jannin	INRIA, France
Leo Joskowicz	Hebrew University of Jerusalem, Israel
Ioannis Kakadiaris	University of Houston, USA
Rasmus Larsen	Technical University of Denmark
Anant Madabhushi	Rutgers University, USA
Frederik Maes	University of Leuven, Belgium
Kensaku Mori	Nagoya University, Japan
Parvin Mousavi	Queen's University, Canada
Nassir Navab	Technical University of Munich, Germany
Poul Nielsen	University of Auckland, New Zealand
Wiro Niessen	Erasmus Medical College, The Netherlands
Sebastien Ourselin	University College London, UK
Nikos Paragios	Ecole Central Paris, France
Xavier Pennec	INRIA, France
Josien Pluim	Utrecht University, The Netherlands
Daniel Rueckert	Imperial College London, UK
Ichiro Sakuma	University of Tokyo, Japan
Tim Salcudean	University of British Colombia, Canada
Julia A. Schnabel	University of Oxford, UK
Dinggang Shen	University of North Carolina, USA
John G. Sled	University of Toronto, Canada

Martin Styner	University of North Carolina, USA
Russell Taylor	Johns Hopkins University, USA
Jocelyne Troccaz	CNRS, France
Aaron Ward	University of Western Ontario, Canada
Simon Warfield	Harvard University, USA
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
Ziv Yaniv	National Children's Hospital, Washington DC, USA

Local Organizing Committee

Janette Wallace	Robarts Research Institute, London, Canada
Johanne Guilemette	Robarts Research Institute, London, Canada
Debbie Lilley	Robarts Research Institute, London, Canada
Jade Orkin-Fenster	Robarts Research Institute, London, Canada
Jackie Williams	Robarts Research Institute, London, Canada
Chris Wedlake	Robarts Research Institute, London, Canada
Shuo Li	GE Medical Systems/University of Western Ontario, London, Canada
Perry Radau	Sunnybrook Research Institute, Toronto, Canada
Raphael Ronen	Sunnybrook Research Institute, Toronto, Canada

Reviewers

Acar, Burak	Audette, Michel
Achterberg, Hakim	Avants, Brian
Acosta-Tamayo, Oscar	Avila, Lisa
Adluru, Nagesh	Awate, Suyash
Aja-Fernández, Santiago	Axel, Leon
Alexander, Daniel	Babalola, Kolawole
Aljabar, Paul	Bach Cuadra, Meritxell
Allain, Baptiste	Baghani, Ali
Amini, Amir	Baka, Nora
An, Jungha	Balicki, Marcin
Arbel, Tal	Baloch, Sajjad
Aribisala, Benjamin	Barbu, Adrian
Ashburner, John	Barmpoutis, Angelos
Astley, Sue	Barratt, Dean
Atasoy, Selen	Batchelor, Philip

Batmanghelich, Nematollah
Baumann, Michael
Baust, Maximilian
Bazin, Pierre-Louis
Beg, Mirza Faisal
Beichel, Reinhard
Bellec, Pierre
Bello, Fernando
Berger, Marie-Odile
Bergtholdt, Martin
Beuthien, Björn
Bhotika, Rahul
Biesdorf, Andreas
Bloch, Isabelle
Blum, Tobias
Boctor, Emad
Boisvert, Jonathan
Bosch, Johan
Bossa, Matias Nicolas
Boucher, Maxime
Bouix, Sylvain
Boukerroui, Djamal
Bourgeat, Pierrick
Brady, Mike
Brock, Kristy
Brost, Alexander
Brun, Caroline
Bullitt, Elizabeth
Buonaccorsi, Giovanni
Burgert, Oliver
Burlina, Philippe
Burschka, Darius
Caan, Matthan
Cahill, Nathan
Camara, Oscar
Camp, Jon
Cardenes, Ruben
Cardoso, Manuel Jorge
Castro-González, Carlos
Cattin, Philippe C.
Cetingul, Hasan Ertan
Chakravarty, M. Mallar
Chen, Elvis C. S.
Chen, Terrence
Chen, Thomas Kuiran
Chen, Ting
Chen, Xinjian
Cheng, Jian
Cheriet, Farida
Chien, Aichi
Chintalapani, Gouthami
Chinzei, Kiyoyuki
Chitphakdithai, Nicha
Chittajallu, Deepak Roy
Chou, Yiyu
Chowdhury, Ananda
Christensen, Gary
Cinquin, Philippe
Ciuciu, Philippe
Clarkson, Matthew
Clouchoux, Cédric
Colliot, Olivier
Combès, Benoît
Commowick, Olivier
Cootes, Tim
Corso, Jason
Cotin, Stephane
Coulon, Olivier
Coupe, Pierrick
Criminisi, Antonio
Crum, William
Cuingnet, Remi
Daga, Pankaj
Dahl, Anders L.
Dalal, Sandeep
Daoud, Mohammad
Darkner, Sune
Darvann, Tron
Dauguet, Julien
Dawant, Benoit
de Boer, Renske
De Craene, Mathieu
Dehghan, Ehsan
Deligianni, Fani
Demiralp, Cagatay
Demirci, Stefanie
Dequidt, Jeremie
Deriche, Rachid
Descoteaux, Maxime
Desvignes, Michel

Dijkstra, Jouke
DiMaio, Simon
Doignon, Christophe
Dong, Bin
Doorly, Denis
Douiri, Abdel
Dowling, Jason
Drangova, Maria
Drechsler, Klaus
Duchateau, Nicolas
Duchesne, Simon
Dufour, Alexandre
Duriez, Christian
Durrleman, Stanley
Ebrahimi, Mehran
Ehman, Richard
Ehrhardt, Jan
El-Baz, Ayman
Elhawary, Haytham
Ellis, Randy
Engelke, Klaus
Enquobahrie, Andinet
Erdt, Marius
Eskandari, Hani
Eskildsen, Simon
Essert, Caroline
Fadili, Jalal
Fallavollita, Pascal
Fan, Yong
Farag, Aly
Farag, Amal
Faure, François
Fedorov, Andriy
Fei, Baowei
Fenster, Aaron
Feulner, Johannes
Figl, Michael
Fillard, Pierre
Fischer, Gregory
Fitzpatrick, J. Michael
Fleig, Oliver
Florack, Luc
Fonov, Vladimir
Foroughi, Pezhman
Fouard, Celine
Freiman, Moti
Freysinger, Wolfgang
Fridman, Yoni
Funka-Lea, Gareth
Gangloff, Jacques
Gao, Wei
Gao, Yi
Garcia-Lorenzo, Daniel
Garvin, Mona
Gee, Andrew
Gee, James
Geng, Xiujuan
Gerig, Guido
Gholipour, Ali
Giannarou, Stamatia
Gibaud, Bernard
Gibson, Eli
Gilles, Benjamin
Gilson, Wesley
Giusti, Alessandro
Glocker, Ben
Gobbi, David
Goh, Alvina
Goksel, Orcun
Gonzalez Ballester, Miguel Angel
Gooding, Mark
Goodlett, Casey
Graham, Jim
Grau, Vicente
Greenspan, Hayit
Grova, Christophe
Guetter, Christoph
Guevara, Pamela
Guizard, Nicolas
Habas, Piotr
Haegelen, Claire
Hager, Gregory D.
Hahn, Horst
Haider, Clifton
Hamm, Jihun
Han, Xiao
Haneishi, Hideaki
Hans, Arne
Hansen, Michael Sass
Hanson, Dennis

Harders, Matthias
Hastreiter, Peter
Hatt, Chuck
Haynor, David
He, Huiguang
He, Yong
Hedjazi Moghari, Mehdi
Hefny, Mohamed
Heimann, Tobias
Heinrich, Mattias Paul
Heldmann, Stefan
Hellier, Pierre
Heng, Pheng Ann
Hernandez, Monica
Higgins, William
Ho, Harvey
Hoffmann, Kenneth
Hogeweg, Laurens
Hollis, Ralph
Hornegger, Joachim
Hu, Mingxing
Hu, Zhihong
Huang, Junzhou
Huang, Wei
Huang, Xiaolei
Huisman, Henkjan
Hyde, Damon
Idier, Jerome
Iglesias, Juan Eugenio
Imani, Farhad
Ionasec, Razvan
Jagadeesan, Jayender
Jahanshad, Neda
Janoos, Firdaus
Jerebko, Anna
Jia, Hongjun
Jiang, Tianzi
Jiang, Yifeng
Jolly, Marie-Pierre
Jomier, Julien
Joshi, Anand
Joshi, Sarang
Jurrus, Elizabeth
Kårsnäs, Andreas
Kabus, Sven
Kadoury, Samuel
Kainmueller, Dagmar
Kamen, Ali
Kapur, Tina
Karamalis, Athanasios
Karemore, Gopal
Karssemeijer, Nico
Kaynig, Verena
Kazanzides, Peter
Keeve, Erwin
Keil, Andreas
Kerrien, Erwan
Khan, Ali R.
Khurd, Parmeshwar
Kikinis, Ron
Kim, Kio
Kim, Minjeong
King, Andrew
Kiraly, Atilla
Kitasaka, Takayuki
Klein, Stefan
Klinder, Tobias
Koikkalainen, Juha
Konukoglu, Ender
Kowal, Jens
Kruggel, Frithjof
Krupa, Alexandre
Kumar, Rajesh
Kurkure, Uday
Kuroda, Yoshihiro
Kutter, Oliver
Kwok, Ka-Wai
Kybic, Jan
Lacefield, James
Laine, Andrew
Landman, Bennett
Lango, Thomas
Langs, Georg
Lapeer, Rudy
Laporte, Catherine
Lashkari, Danial
Law, Max W.K.
Le, Yen
Lee, Junghoon
Lee, Su-Lin

Lee, Tim
Lekadir, Karim
Lenglet, Christophe
Lensu, Lasse
Leow, Alex
Lepore, Natasha
Lerch, Jason
Leung, Kelvin
Li, Chunming
Li, Fuhai
Li, Gang
Li, Kaiming
Li, Kang
Li, Ming
Li, Ruijiang
Li, Shuo
Li, Yang
Liang, Jianming
Liao, Hongen
Liao, Rui
Liao, Shu
Lindseth, Frank
Linguraru, Marius George
Linte, Cristian
Liu, Chi
Liu, Huafeng
Liu, Jiamin
Liu, Sheena
Liu, Tianming
Liu, Xiaofeng
Lo, Pechin
Loeckx, Dirk
Loew, Murray
Lori, Nicolas
Lu, Chao
Lu, Le
Luboz, Vincent
Lui, Lok Ming
Ma, YingLiang
Mørup, Morten
Müller, Henning
Machann, Juergen
Machiraju, Raghu
Maddah, Mahnaz
Madore, Bruno
Magee, Derek
Magill, John
Maier-Hein, Lena
Makram-Ebeid, Sherif
Malandain, Gregoire
Manduca, Armando
Mangin, Jean-Francois
Mani, Meenakshi
Manjon, Jose V.
Manniesing, Rashindra
Mansi, Tommaso
Manzke, Robert
Marchal, Maud
Marsland, Stephen
Martí, Robert
Martin-Fernandez, Marcos
Masamune, Ken
Masutani, Yoshitaka
Mattes, Julian
Maurel, Pierre
McClelland, Jamie
Melbourne, Andrew
Menze, Bjoern
Metaxas, Dimitris
Metz, Coert
Meyer, Chuck
Michel, Fabrice
Miga, Michael
Miller, James
Miller, Karol
Mirota, Daniel
Modat, Marc
Modersitzki, Jan
Mohamed, Ashraf
Montillo, Albert
Moore, John
Moradi, Mehdi
Mountney, Peter
Murgasova, Maria
Mylonas, George
Naish, Michael
Nakamura, Ryoichi
Nash, Martyn
Nassiri-Avanaki, Mohammad-Reza
Nichols, Thomas

Nicolau, Stéphane
Niemeijer, Meindert
Niethammer, Marc
Noël, Peter
Noble, Alison
Nolte, Lutz
Ocegueda, Omar
Oda, Masahiro
O'Donnell, Lauren
O'Donnell, Thomas
Ogier, Arnaud
Oguz, Ipek
Okada, Toshiyuki
Olabarriaga, Silvia
Oliver, Arnau
Olmos, Salvador
Orihuela-Espina, Felipe
Orkisz, Maciej
Otake, Yoshito
Ou, Yangming
Oubel, Estanislao
Ozarslan, Evren
Pace, Danielle
Padfield, Dirk
Padoy, Nicolas
Palaniappan, Kannappan
Paniagua, Beatriz
Papadakis, Emanuel
Papademetris, Xenios
Parthasarathy, Vijay
Passat, Nicolas
Patel, Rajni
Patriciu, Alexandru
Paul, Perrine
Paulsen, Rasmus
Pauly, Olivier
Pavlidis, Ioannis
Pearlman, Paul
Peitgen, Heinz-Otto
Pekar, Vladimir
Penney, Graeme
Pernus, Franjo
Petersen, Jens
Peyrat, Jean-Marc
Pham, Dzung
Pitiot, Alain
Pizaine, Guillaume
Platel, Bram
Plewes, Donald
Podder, Tarun
Pohl, Kilian Maria
Poignet, Philippe
Poot, Dirk
Poynton, Clare
Prager, Richard
Prastawa, Marcel
Pratt, Philip
Prima, Sylvain
Prince, Jerry
Punithakumar, Kumaradevan
Qazi, Arish A.
Qian, Xiaoning
Qian, Zhen
Qin, Lei
Quellec, Gwenole
Radeva, Petia
Rajagopal, Vijayaraghavan
Ramamurthi, Krishnakumar
Ramezani, Mahdi
Rasoulian, Abtin
Rathi, Yogesh
Reinertsen, Ingerid
Rettmann, Maryam
Reyes, Mauricio
Reyes-Aldasoro, Constantino
Rhode, Kawal
Ribbens, Annemie
Risholm, Petter
Risser, Laurent
Rit, Simon
Rittscher, Jens
Rivaz, Hassan
Riviere, Cameron
Robb, Richard A.
Roche, Alexis
Rohkohl, Christopher
Rohlfing, Torsten
Rohling, Robert
Rohr, Karl
Ross, James

Rousseau, François
Russakoff, Daniel
Sörensen, Lauge
Sabuncu, Mert Rory
Sadeghi Naini, Ali
Saha, Punam Kumar
Salvado, Olivier
Sanchez, Clara
Savadjiev, Peter
Savinaud, Mickaël
Schaller, Christian
Scherrer, Benoit
Schultz, Thomas
Schweikard, Achim
Sermesant, Maxime
Shah, Mohak
Shah, Shishir
Shamir, Reuben R.
Shams, Ramtin
Shen, Li
Shi, Feng
Shi, Pengcheng
Shi, Yonggang
Shi, Yundi
Shimizu, Akinobu
Siddiqi, Kaleem
Siewerdsen, Jeffrey
Simaan, Nabil
Simpson, Amber
Singh, Nikhil
Sinkus, Ralph
Sjöstrand, Karl
Slabaugh, Greg
Slagmolen, Pieter
Smal, Ihor
Smeets, Dirk
Soler, Luc
Song, Sang-Eun
Song, Xubo
Sotiras, Aristeidis
Sporring, Jon
Staib, Lawrence
Staring, Marius
Stewart, James
Stoianovici, Dan
Stoyanov, Danail
Strother, Stephen
Studholme, Colin
Subramanian, Navneeth
Summers, Ronald
Sun, Bo
Sundar, Hari
Suzuki, Kenji
Syeda-Mahmood, Tanveer
Szczerba, Dominik
Szekely, Gabor
Szilagyi, Laszlo
Sznitman, Raphael
Tahmasebi, Amir
Tanner, Christine
Tao, Xiaodong
Tardif, Christine Lucas
Tasdizen, Tolga
Taylor, Zeike
Thévenaz, Philippe
Thiran, Jean-Philippe
Thiriet, Marc
Thirion, Bertrand
Tiwari, Pallavi
Todd Pokropek, Andrew
Todd, Catherine
Toews, Matthew
Tokuda, Junichi
Toussaint, Nicolas
Tristán-Vega, Antonio
Tsekos, Nikolaos V.
Tsin, Yanghai
Unal, Gozde
Ungi, Tamas
Vaillant, Regis
van Assen, Hans
van der Lijn, Fedde
van Ginneken, Bram
Van Leemput, Koen
van Rikxoort, Eva
van Stralen, Marijn
Vandergheynst, Pierre
Vannier, Michael
Varoquaux, Gael
Vegas-Sánchez-Ferrero, Gonzalo

Venkataraman, Archana
Vercauteren, Tom
Verma, Ragini
Vermandel, Maximilien
Viergever, Max
Vignon, Francois
Villard, Pierre-Frederic
Visentini-Scarzanella, Marco
Vitanovski, Dime
Vitiello, Valentina
von Berg, Jens
Voros, Sandrine
Vos, Pieter
Vosburgh, Kirby
Votta, Emiliano
Vrooman, Henri
Vrtovec, Tomaz
Wachinger, Christian
Waechter-Stehle, Irina
Wahle, Andreas
Wang, Defeng
Wang, Fei
Wang, Hongzhi
Wang, Lejing
Wang, Liansheng
Wang, Peng
Wang, Qian
Wang, Song
Wang, Yang
Wang, Ying
Wassermann, Demian
Weber, Stefan
Weese, Jürgen
Whitaker, Ross
Wiest-Daessle, Nicolas
Wiles, Andrew
Wirtz, Stefan
Wittek, Adam
Wolf, Ivo
Wollny, Gert
Wolz, Robin
Wörz, Stefan
Wu, Guorong
Wu, John Jue
Wu, Xiaodong
Xi, Zhao
Xiao, Gaoyu
Xie, Hua
Xie, Jun, Jun
Xiong, Guanglei
Xu, Jun
Xu, Sheng
Xue, Zhong
Yalamanchili, Raja
Yan, Pingkun
Yang, Xiaoyun
Yao, Jianhua
Yap, Pew-Thian
Ye, Dong Hye
Yeo, B.T. Thomas
Yin, Youbing
Yin, Zhaozheng
Yoo, Terry
Yoshida, Hiro
Young, Stewart
Yushkevich, Paul
Zacharia, Eleni
Zahiri Azar, Reza
Zeng, Wei
Zhan, Liang
Zhang, Daoqiang
Zhang, Dongping
Zhang, Hui
Zhao, Fei
Zheng, Guoyan
Zheng, Yefeng
Zheng, Yuanjie
Zhong, Hua
Zhou, Jinghao
Zhou, Luping
Zhou, S. Kevin
Zhou, X. Sean
Zhu, Hongtu
Zhu, Ying
Zikic, Darko
Zion, Tse
Zwiggelaar, Reyer

Awards Presented at MICCAI 2010, Beijing

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 2010 Enduring Impact Award was presented to Russ Taylor, Johns Hopkins University, USA.

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 2010, fellowships were awarded to:

James S. Duncan (Yale University, USA)
Stephen M Pizer (University of North Carolina, USA)
Jocelyne Troccaz (CNRS, France)

MedIA-MICCAI Prizes (Split decision)
Jihun Hamm, for the article entitled: "GRAM: A Framework for Geodesic Registration on Anatomical Manifolds," co-authored by: Jihun Hamm, Dong Hye Ye, Ragini Verma, Christos Davatzikos

Samuel Gerber, for the article entitled: "Manifold Modeling for Brain Population Analysis," co-authored by: Tolga Tasdizen, P. Thomas Fletcher, Sarang Joshi, Ross Whitaker

Best Paper in Computer-Assisted Intervention Systems and Medical Robotics, Sponsored by Intuitive Surgical Inc.
Rogerio Richa, for the article entitled: "Robust 3D Visual Tracking for Robotic-Assisted Cardiac Interventions," co-authored by: Rogerio Richa, Antonio P. L. Bo, and Philippe Poignet

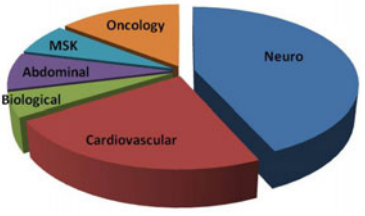
MICCAI Young Scientist Awards
The Young Scientist Awards are stimulation prizes awarded for the best first authors of MICCAI contributions in distinct subject areas. The nominees had to be full-time students at a recognized university at, or within, two years prior to submission. The 2010 MICCAI Young Scientist Awards were presented in the following categories to:

Instrument and Patient Localization and Tracking	**Ehsan Dehghan** "Prostate Brachytherapy Seed Reconstruction Using C-Arm Rotation Measurement and Motion Compensation"

Image Reconstruction and Restoration	**Junzhou Huang** "Efficient MR Image Reconstruction for Compressed MR Imaging"
Modelling and Simulation	**Saša Grbić** "Complete Valvular Heart Apparatus Model from 4D Cardiac CT"
Quantitative Image Analysis	**Rémi Cuingnet** "Spatially Regularized SVM for the Detection of Brain Areas Associated with Stroke Outcome"
Functional and Diffusion-Weighted MRI	**Anthony J. Sherbondy** "MicroTrack: An Algorithm for Concurrent Projectome and Microstructure Estimation"

Accepted MICCAI 2011 Papers

by Clinical Theme:



by Technical Theme:

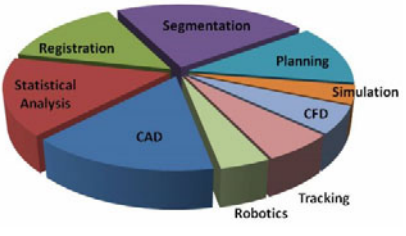


Table of Contents – Part I

Robotics, Localization and Tracking and Visualization I

Robotics, Localization and Tracking and Visualization II

Robotics, Localization and Tracking and Visualization III

Planning and Image Guidance

Physical Modeling and Simulation I

Physical Modeling and Simulation II

Motion Modeling and Compensation

Segmentation and Tracking in Biological Images

Table of Contents – Part II

Diffusion Weighted Imaging I

Diffusion Weighted Imaging II

fMRI

Statistical Analysis and Shape Modelling I

Statistical Analysis and Shape Modelling II

Registration I

Registration II

Table of Contents – Part III

Computer Aided Diagnosis and Machine Learning I

Computer Aided Diagnosis and Machine Learning II

Computer Aided Diagnosis and Machine Learning III

Segmentation I

Segmentation II

Segmentation III

Unified Detection and Tracking in Retinal Microsurgery

Raphael Sznitman, Anasuya Basu, Rogerio Richa, Jim Handa, Peter Gehlbach, Russell H. Taylor, Bruno Jedynak, and Gregory D. Hager

Johns Hopkins Univesity
3400 N. Charles Str., Baltimore, USA
{sznitman,abasu8,richa,rht,bruno.jedynak,hager}@jhu.edu,
{jthanda,pgelbac}@jhmi.edu

Abstract. Traditionally, tool tracking involves two subtasks: (i) detecting the tool in the initial image in which it appears, and (ii) predicting and refining the configuration of the detected tool in subsequent images. With retinal microsurgery in mind, we propose a unified tool detection and tracking framework, removing the need for two separate systems. The basis of our approach is to treat both detection and tracking as a sequential entropy minimization problem, where the goal is to determine the parameters describing a surgical tool in each frame. The resulting framework is capable of both detecting and tracking in situations where the tool enters and leaves the field of view regularly. We demonstrate the benefits of this method in the context of retinal tool tracking. Through extensive experimentation on a phantom eye, we show that this method provides efficient and robust tool tracking and detection.

1 Introduction

Surgical tool tracking has recently established itself as part of the computer assisted intervention community [1,2]. Ophthalmic microsurgery is an emerging research field, and the ability to detect and track retinal tools is very important for automatic vein cannulation, retinal modeling, precise retinal image mosaicking and other microretinal surgery applications. Informally, the objective of visual tracking is to provide an accurate estimate of the parameters or configuration of a tool across time. A general solution involves two subtasks: (i) detecting the tool in the initial image in which it appears, and (ii) predicting and refining (*i.e.* tracking) the configuration of the detected tool in subsequent images [3].

Indeed, few accurate tool detection and tracking systems exists in the context of retinal microsurgery. In [4] tracking is performed frame by frame by using color appearance and a 3D tool model, rendering tracking relatively slow. Similarly, [5] casts the retinal tool tracking as a linear programing problem, achieving good accuracy but also suffering from high computational cost. Gradient-based color tracking in retinal image sequences was also proposed in [2,6], but is highly depend on good initialization. More generally within the context of computer

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 1–8, 2011.

vision, combining detection and tracking remains a difficult problem when an object enters and leaves the field of view frequently.

We propose an algorithmic framework that combines tool detection and tracking in an information theoretic setting. Instead of treating detection and tracking as separate tasks, we cast both as a single sequential entropy minimization problem, treating tool detection and sequential localization estimation as one. Consequently, the system presented here consists of a single process, as opposed to two, and hence reduces the need for careful initializations and parameter tuning. Our solution solves detection and tracking as a density estimation problem by using an *Active Testing* (AT) [7,8] optimization strategy. In practice, we demonstrate that our approach provides a feasible and automatic solution to tool detection and tracking, without the need of accurate motion models. While we have chosen to show this method in the context of retinal tool tracking, we believe it is general enough for kinematic tools in a number of other surgical settings.

The remainder of the paper is organized as follows: in Sec. 2 we first describe tracking as a Bayesian sequential estimation problem and how our approach embodies this structure. We then describe how the AT paradigm is used to locate tools in Sec. 3. In Sec. 4, we perform extensive experimentation to validate our approach. Finally, we conclude with some closing remarks in Sec. 5.

2 Tool Tracking

In what follows, we parametrize the surgical tool by (1) the location on the boundary of the image where the tool enters, (2) the angle the tool makes with the image boundary and (3) the length of tool (see Fig. 1(a)). Note that this choice simply reflects known constraints for this application. More formally, the tool configuration space is defined as $Y = (y_1, y_2, y_3) \in \mathcal{S}_1$ where, $\mathcal{S}_1 = [0, P] \times [-\pi/2, \pi/2] \times [\delta, D]$, where P is the length of the perimeter of the image in pixels, δ is a minimal length the tool must protrude before being considered visible and D is the diagonal image length. To model the tool leaving the field of view, we allow a special token $\mathcal{S}_0$ indicating that the tool is not visible (*i.e.* $Y \notin \mathcal{S}_1$). Thus, $Y \in \mathcal{S} = \mathcal{S}_0 \cup \mathcal{S}_1$.

We cast the tracking problem in a Bayesian sequential estimation fashion. That is, at time t, we consider a random variable Y^t that must be inferred given the image sequence observed up to that time instance, $\mathcal{I}^t = (I^0, \ldots, I^t)$. Using a standard Hidden Markov Model assumption, we formulate this as a Bayesian filtering problem [9]. That is, we specify a prior distribution on Y, denoted $P(Y^0)$, a tool dynamics model, $P(Y^t|Y^{t-1})$, and will sequentially estimate the new tool parameters given the observed images $\mathcal{I}^t$.

Active Testing Filtering: In order to compute the posterior distribution of Y given the history of observations, $P(Y^t|\mathcal{I}^t)$, we make use of the AT strategy. AT is a stochastic optimization technique used for parameter estimation. When trying to determine a set of parameters, $Y = (y_1, \ldots, y_n)$, this technique requires a prior

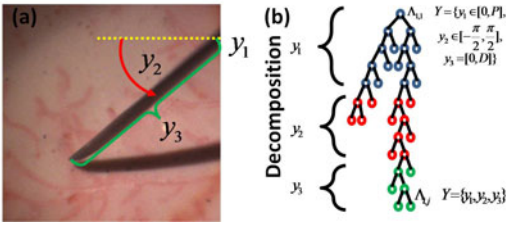


Fig. 1. Parametrization of retinal tool: (a) Visual representation of the tool parametrization (b) Decomposition of tool pose space

probability distribution on the parameters, which in our case will be $P(Y^t|\mathcal{I}^{t-1})$, a set of questions, $\mathcal{X}$, pertaining to the parameters and a parametrization of the parameter space. The main idea is to use the available questions to make the distribution of Y evolve such that it is peaked on the correct parameters, (*i.e.* tool location and pose). AT automatically selects which questions to ask by using an entropic loss function. Algorithmically, this consists in: (i) select a question (*i.e.* an image functional) and subset of the parameter space, (ii) apply this question to the subset selected, (iii) update the probability distribution of the parameters, (iv) select the next question and subset pair that maximizes the mutual information gain, (v) repeat from (ii) until the entropy of the distribution is low. For a more thorough review of AT, see [7,8].

Given this, we propose a general Active Testing Filter (ATF) (see Alg. 1). The user initially provides, some dynamics model, $P(Y^t|Y^{t-1})$ and an initial prior $P(Y^0)$. Then, for an image in the sequence, first compute $P(Y^t|\mathcal{I}^{t-1})$ (line: 3) by using the provided dynamics model. Then treat $P(Y^t|\mathcal{I}^{t-1})$ as an initial prior for the AT localizer (line: 4). This process is then repeated for all images in the sequence

In the following section, we specify the required question set and space parametrization.

Algorithm 1. Active Testing Filtering ($\mathcal{I} = \{I^1, \ldots, I^T\}$)

1: Initialize: $P(Y^0), P(Y^t|Y^{t-1})$
2: **for all** $t = 1, \ldots, T$ **do**
3: $\quad P(Y^t|\mathcal{I}^{t-1}) = \int P(Y^t|Y^{t-1})P(Y^{t-1}|\mathcal{I}^{t-1})dY^{t-1}$
4: $\quad P(Y^t|\mathcal{I}^t) = ActiveTesting(I^t, P(Y^t|\mathcal{I}^{t-1}))$
5: **end for**

3 Active Testing for Tool Localization

To use the AT framework, one must provide a question set and specify a partitioning of the search space. We now describe these specifications.

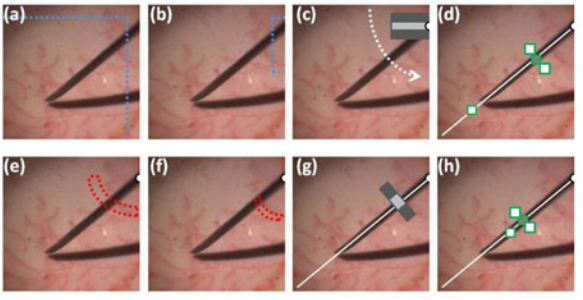


Fig. 2. Question Set. Example images with each image designating the queried region. See text for details on these queries.

Space Representation: Let Λ denote a binary decomposition of the space $\mathcal{S}_1$. That is, Λ is a tree of sets, $\Lambda = \{\Lambda_{i,j}, i = 0, \ldots, M, j = 0, \ldots, 2^i - 1\}$ (see Fig. 1(b)). The root of tree is $\Lambda_{0,0} = \mathcal{S}_1$. The decomposition of the tree is performed by splitting one coordinate of Y at a time, until a desired resolution, at which point we repeat the procedure for another coordinate (*e.g.* split y_1, then y_2 and so on). That is, the partition forms conditional subtrees: the first subtree decomposes the first parameter, while the second subtree decomposes the second parameter conditioned on the first, and the last subtree partitions the final parameter conditioned on the two first parameters. The subtrees are color coded in Fig. 1(b). It is easy to show that any level of the tree forms a covering on $\mathcal{S}_1$, *i.e.* , $\mathcal{S}_1 = \Lambda_{0,0} = \bigcup_{j=0}^{2^i-1} \Lambda_{i,j}$.

Query Set: To determine the target configuration, a set of questions or tests, $\mathcal{X}$, pertaining to the target must be provided. This consists in associating each node $\Lambda_{i,j}$ with a set of pose-indexed queries that provide information on whether or not the target has configuration contained within the node at hand. For each node $\Lambda_{i,j}$, we define a set of questions, $X_{i,j} = \{X_{i,j}^1, \ldots, X_{i,j}^K\}$ where $X_{i,j}^k$ is some image functional; $X : I \mapsto \mathbb{R}$. Finally, all questions are assumed to have homogeneous responses such that,

$$P(X_{i,j}^k = x | Y = y) = \begin{cases} f_o^k(x) & \text{if } y \in \Lambda_{i,j} \\ f_b^k(x) & \text{if } y \notin \Lambda_{i,j} \end{cases} \tag{1}$$

where f_o^k and f_b^k are two distributions of responses, corresponding to the case where the tool configuration is in the space queried, and when it is not. Since the response, x to the query $X_{i,j}^k$ will be in $\mathbb{R}$, both f_o^k and f_b^k will be modeled as Gaussians. The parameters of these distributions (*i.e.* μ, σ^2) are learned from separate training images representative of test sequences.

We now specify the query set, $\mathcal{X}$. In total, we form six queries. That is for each node in Λ, we have $X_{i,j} = \{X_{i,j}^1, \ldots, X_{i,j}^6\}$. We let $X_{i,j}^1$ count the proportion of pixels deemed tool like, by using a color model learned from data, over a region determined by the interval of y_1 in $\Lambda_{i,j}$. An example is shown in Fig. 2(a)-(b),

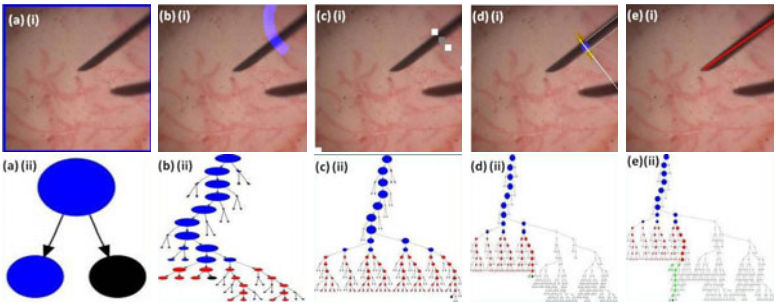


Fig. 3. Active Testing Iterations. Each image pair (top and bottom row) show a query being evaluated and the corresponding state of the AT tree at that point in time, respectively. See text for details.

where each image shows a different $\Lambda_{i,j}$ being queried by $X^1_{i,j}$. $X^2_{i,j}$ also queries a part of the y_1 interval by applying a small template to the image boundary centered on y_1, and compute the minimum template score when varying the template orientation (see Fig 2(c)).

In order to estimate y_2, we form $X^3_{i,j}$ and $X^4_{i,j}$. Given a fixed tool boundary point, $X^3_{i,j}$, computes the proportion of tool pixels (as in $X^1_{i,j}$) which are present in an arc defined by the pose interval of y_2 in $\Lambda_{i,j}$ (see Fig. 2(e)-(f) for two examples). $X^4_{i,j}$ again applies a template match perpendicular to the average angle in the interval of y_2 in $\Lambda_{i,j}$ (see Fig. 2(g)).

Estimating y_3 is achieved by queries $X^5_{i,j}$ and $X^6_{i,j}$. At this point, both the tool boundary point and the angle are assumed to be fixed. Estimating if the length of the tool is in the interval of y_3 in $\Lambda_{i,j}$ is done by computing the difference in average pixel intensities at the upper bound and lower bounds of y_3 in $\Lambda_{i,j}$ (see Fig. 2(d)-(h)). Finally, $X^6_{i,j}$ performs a template match according to the parameters specified in $\Lambda_{i,j}$.

4 Experiments

We choose the prior of Y to be uninformative and let $P(Y^0) = P(\{y_1 \in [0, P], y_2 \in [-\pi/2, \pi/2], y_3 \in [0, P]\}) = 1/2$, indicating equal likelihood that tool is or is not in the image. Here, we use a simple linear dynamic model of the form, $Y^{t+1} = AY^t + \mathcal{N}(0, \alpha)$, where A is the dynamics transition matrix. In the experiments that follow, A is augmented to allow velocity estimates to be compounded in the new prior. Given that we know that the tool will enter and leave the field of view often, we expect the dynamics model to be violated regularly. While this may induce inappropriate priors $P(Y^t|\mathcal{I}^{t-1})$ at each step, the AT framework will correct for this automatically. Note that, the AT model is trained on a separate and representative training set of 50 images, where the tool pose has been annotated. Also, since we are only interested in tracking the tool when it is in focus and the depth of field is particularly small on retinal

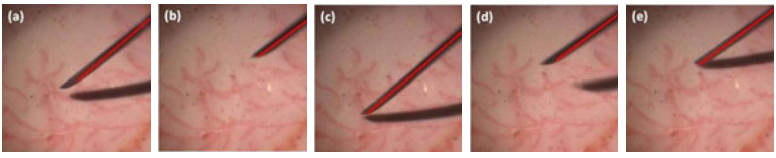


Fig. 4. Snapshots of detection and tracking during an image sequence

microscopes, the tool scale exhibits relatively little variability. Hence, we will assume very limited tool scale changes.

Experimental Setup: Similarly to [4,2], we recorded two video sequences of a retinal tool interacting with a phantom retinal membrane. The sequence consists of 400 frames of size 256×256. In these sequences, a retinal tool (needle) is initially present in the field of view and moves regularly. The tool leaves and re-enters the field of view multiple times throughout these sequences. In images that contain the tool, the locations have been manually annotated for quantitative analysis evaluation. All algorithms were tested on a standard 2.3GHz PC.

To illustrate the AT step in our ATF approach, Fig. 3 shows both the queries asked and the evolution of Λ at selected iterations during this process. The top row shows which query is being evaluated with the queried region highlighted in each image. The bottom row shows the state of Λ at that point. The area of each node is proportional to the mass contained in that pose subset, while the color of each node represents which coordinate is being refined (as in Fig. 1).

Initially, only the root $\Lambda_{0,0}$ exists and a query is evaluated on the entire pose space. Having created children (Fig. 3(a)i-ii), the size of Λ consists of three nodes. After a few queries, the tree has grown and refined itself past the first coordinate y_1, onto y_2 and y_3 (Fig. 3(b)(c)i-ii). Eventually the correct y_2 parameter (Fig. 3(d)i-ii) is located, leading to a valid tool detection (Fig. 3(e)i-ii).

Having detected the tool parameters, we use the dynamics model to compute a new prior that seeds another round of the AT step. As such, no separate mechanism is required to initialize tracking, or filter responses. Fig. 4 shows some selected images of the tool being tracked. When the tool moves beyond the domain of the image, tracking is abandoned, correctly detecting that no tool is present. The object then reappears in the field of view and is automatically reacquired. Notice that when the tool tip and tool shadow approach together, the tool object is still correctly tracked (Fig. 4(c)(e)).

Comparison: To evaluate the performance of the ATF approach, we compared it with two methods: Simple Detect (SD) and Simple Detection and Tracking (SDT). Both these methods are specially tailored using *a priori* knowledge of the appearance of the tool and background in the experimental image sequences.

In SD, all pixel positions are tested to detect the tip of a long shaft. The pixel intensity threshold for separating tool and background was tuned empirically so that no false detections occur. Once it is detected, the tool is tracked using the tool position from the previous image for initialization.

Method	Accuracy Error				Detection			Time (ms)
	y_1	y_2	y_3	Tip	TPR	FPR	Precision	
ATF	**3.51 (0.13)**	**1.79 (0.1)**	**12.67 (0.85)**	**11.51 (0.84)**	**0.839**	6.4×10^{-6}	**0.661**	7.21
AT	5.53 (0.70)	2.41 (0.13)	14.16 (1.02)	13.49 (0.81)	0.811	6.1×10^{-6}	0.615	25.21
SD	84.87 (2.45)	29.47 (0.81)	20.33 (0.94)	15.19 (0.79)	0.778	7.1×10^{-6}	0.547	26.67
SDT	52.58 (7.41)	11.09 (0.92)	21.24 (2.3)	11.64 (0.24)	0.834	$\mathbf{7.4\times10^{-6}}$	0.596	4.8

Fig. 5. Performance Results

In a similar spirit to [2], SDT includes a tracking loop after detection (as performed by SD). The tool tracking loop consists of the following steps. First, for a new image, the displacement of the tool is found by searching along a line perpendicular to the tool. Second, the new tool angle is measured. The maximal angle variation is set to 35 degrees for limiting the computational search cost. Finally, the tool tip is found by searching along the tool shaft for the intersection of the tool and the background (the intersection is determined by thresholding the intensity level using empirically chosen values)[1].

In our comparison we observe three aspects of performance. First, we compute the error in the estimates of each parameter and the tool tip position. This error is computed by using the annotated ground truth. Similarly, we compute statistics which are typically observed in detection and localization tasks. We consider a correct detection to be one which estimates the tool tip location to within 10 pixels of the ground truth. We then compute the *true positive rate* (TPR), the *false positive rate* (FPR) and the precision for each approach. Finally, we report the computational time of each algorithm. Given these performance measures, we compare the following four algorithms: the Active Testing algorithm (AT), both comparison methods described above, SD and SDT, and the ATF algorithm.

Fig. 5 summarizes our experimental results. For the accuracy error, we report the means and standard errors for each tool parameter and tool tip. Notice that in general all the methods proposed provide more or less the same detection accuracy. Naturally, we see that detection is significantly slower than tracking with both AT and SD running much slower than tracking methods, confirming the advantages of tracking strategies over tracking by pure detection approaches.

From the algorithms tested, ATF appears to outperform other methods and yet remains computationally efficient (*i.e.* over 90 fps). In the domain of accuracy, the ATF approach estimates the tool parameters more accurately and consistently. When compared to SDT, ATF is generally capable of determining the configuration of the tool more accurately. This improvement can be attributed to the sequential parameter estimation that the active testing framework conducts. By estimating the first parameter, then the second and so on, each parameter is individually estimated accurately. This is in sharp contrast to the more direct SDT approach which locates the tool tip, and then estimates the necessary parameters.

[1] See project website for additional information.

5 Conclusion

We have proposed a novel approach to tool tracking and detection. By treating both problems in a sequential tool parameter estimation setting, our framework minimizes the entropy of the joint distribution over the tool parameters and the available questions provided. A full detection and tracking algorithm has been outlined, and we have empirically shown that the proposed method is capable of detecting and tracking retinal tools efficiently and robustly in cases where the tool enters and leaves the field of view frequently. As part of our future work, we are looking at extending this framework to tracking multiple targets such that this approach may be applicable in a larger number of settings.

Acknowledgments. Funding for this research was provided in part by NIH Grant R01 EB 007969-01, an unrestricted grant by RPB (Wilmer Eye Institute), internal JHU funds and a research subcontract from Equinox Corporation.

References

1. 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. IJRR 26, 1173–1190 (2007)
2. Dewan, M., Marayong, P., Okamura, A.M., Hager, G.D.: Vision-based assistance for ophthalmic micro-surgery. In: Barillot, C., Haynor, D.R., Hellier, P. (eds.) MICCAI 2004. LNCS, vol. 3217, pp. 49–57. Springer, Heidelberg (2004)
3. Yilmaz, A., Javed, O., Shah, M.: Object tracking: A survey. ACM Computing Surveys 38(4) (2006)
4. Sznitman, R., Rother, D., Handa, J., Gehlbach, P., Hager, G.D., Taylor, R.: Adaptive multispectral illumination for retinal microsurgery. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6363, pp. 465–472. Springer, Heidelberg (2010)
5. Pezzementi, Z., Voros, S., Hager, G.D.: Articulated object tracking by rendering consistent appearance parts. In: IEEE international Conference on Robotics and Automation, pp. 1225–1232. IEEE Press, Piscataway (2009)
6. Burschka, D., Corso, J.J., Dewan, M., Lau, W., Lia, 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(1), 5–26 (2005)
7. Geman, D., Jedynak, B.: An active testing model for tracking roads in satellite images. IEEE TPAMI 18(1), 1–14 (1996)
8. Sznitman, R., Jedynak, B.: Active testing for face detection and localization. IEEE TPAMI 32(10), 1914–1920 (2010)
9. Thrun, S., Burgard, W., Fox, D.: Probabilistic Robotics (Intelligent Robotics and Autonomous Agents). The MIT Press, Cambridge (2005)

Graph-Based Geometric-Iconic Guide-Wire Tracking

Nicolas Honnorat[1,2,3], Régis Vaillant[1], and Nikos Paragios[2,3,⋆]

[1] General Electric Healthcare, Buc, France
[2] Laboratoire MAS, Ecole Centrale Paris, Châtenay-Malabry, France
[3] Equipe GALEN, INRIA Saclay-Ile-de-France, Orsay, France

Abstract. In this paper we introduce a novel hybrid graph-based approach for Guide-wire tracking. The image support is captured by steerable filters and improved through tensor voting. Then, a graphical model is considered that represents guide-wire extraction/tracking through a B-spline control-point model. Points with strong geometric interest (landmarks) are automatically determined and anchored to such a representation. Tracking is then performed through discrete MRFs that optimize the spatio-temporal positions of the control points while establishing landmark temporal correspondences. Promising results demonstrate the potentials of our method.

Keywords: Guide-wire, tracking, MRF, simultaneous geometric-iconic registration, landmarks matching.

1 Introduction

Several works tried to address the tracking of the guide-wires (GW) used during cardiac angioplasty interventions for positioning surgical devices into the artery to be cured, because a reliable localization of these wires would be valuable for the intervention monitoring and assistance. This task inherits unfortunately important technical and theoretical challenges. Although the tip of these wires is usually made of a more absorbent material, the main part of these wires is often barely detectable for conventional curvilinear operators [2]. The detection literature provides numerous alternatives, such as Hessian regularized by coherence enhancing diffusion [1], Vesselness [6] and phase congruency [12]. Boosted detectors combining Haar features or steerable filters and image statistics have been used recently [14,7] but these better detectors are more complex or time consuming. The second challenge is related to the large and unforeseeable motion of the GW, due to the combination of cardiac/respiratory motions and patient displacement. Their accumulation cannot be easily captured using conventional prediction mechanisms. Most of the existing tracking approaches represent GW

⋆ This work was supported by ANRT (grant 1062/2008), GE Healthcare and the European Research Council Starting Grant Diocles (ERC-STG-259112). The authors thank T. Hauke Heibel, the reviewers and S. Bernhardt for their help and comments.

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 9–16, 2011.

with B-splines which optimal displacement is found through optimization. In [1] Powell's direction set method was used for tracking rigid GW tips while [14] proposed a method carrying out a rigid alignment followed by a non-rigid registration displacing control-points.

Other GW tracking methods can be associated with well-known contour tracking algorithms. The B-spline snakes have thus been exploited in [12] with an external energy driven from X-ray phase congruency, an internal energy favoring low curvature and a length prior. Methods based on dynamic programming and graphical models were also investigated. The space of possible displacements was for instance discretized and optimal ones were determined using dynamic programming. The CONDENSATION algorithm [8] was applied successfully to tracking contours represented with B-splines. Nevertheless, the space of contour localizations was reduced to a shape space. A parametrized curve tracker based on a Markov Random Field (MRF) framework was recently proposed in [6] and improved in [11]. The main advantage of these approaches is that they achieve near-optimal solution at reasonable computational complexity.

Our approach lies within this scope. First, we address the low signal-to-noise ratio by using steerable filters designed for ridge detection [9] that are enhanced by tensor voting [4,3]. Toward coping with important displacements, we introduce a unified framework which combines an iconic B-spline tracking [6] with a landmark matching approach [13]. We improve the iconic part by using a data term more robust with respect to false GW detections and a prior penalizing also wires rotations. The resulting formulation recovers both landmarks position and the GW displacements through the use of image support and spatio-temporal consistency. This is achieved through a two-part graph acting on the two sets of variables which are coupled in a rigorous mathematical manner. The overall formulation is solved using efficient linear programming.

2 Hybrid Curve Tracking

Iconic tracking methods allow to track non-rigid structures precisely but can be sensitive to local minima (due to erroneous image support) and, depending on the optimization schema, can fail to account for large displacements.

On the contrary, geometric methods rely on the matching of salient structures and can therefore deal with large displacements efficiently. On the other hand, deformations are sparse and incomplete since information for the motion/displacement of the structure between landmark points is not recovered.

Tracking GW during cardiac interventions involves large displacements and non-rigid deformations. Hence we propose a unified discrete framework like [13] combining a variant of the iconic tracking of [6] with a geometric tracking relying on landmarks extracted along the GW. This framework also model exactly the interaction between these two parts. Without loss of generality, let us first consider a representation of the wire with N control-points $\mathbf{c}(i)$ and a set of M landmarks $\mathbf{p}(j)$ extracted along the wire. These variables evolve in time towards representing the structure of interest, or $\mathcal{X}(t) = \{\mathbf{c}(i;t), \mathbf{p}(j;t)\}$.

In terms of support, let us consider a vectorial feature image where measurements of the strength $g(\mathbf{x};t)$ and the orientation $\theta(\mathbf{x};t)$ of the GW are combined: $\mathcal{F}(t) = \{g(\mathbf{x};t)cos(\theta(\mathbf{x};t)), g(\mathbf{x};t)sin(\theta(\mathbf{x};t)\}$. Let us also consider that for the different landmark points $\mathbf{p}(j;t)$, candidate correspondences have been determined at the frame $t+1$ denoted with $\mathcal{P} = \{\mathbf{p}^L(j;m;t+1)\}$, where L is the number of candidates per landmark and $m \in \{1 \dots L\}$. GW tracking is then equivalent to finding the optimal configuration for $\mathcal{X}(t+1)$, given $\mathcal{X}(t)$, $\mathcal{F}(t+1)$ and $\mathbf{p}^L(j;m;t+1)$.

We adopt a unified first order MRF framework combining unary and pairwise potentials involving either the displacements l_i of $\mathbf{c}(i;t)$ or the matching l_j of $\mathbf{p}(j;t)$, originated from data, prior, landmarks and coupling terms:

$$E_{MRF} = \sum_{x \in \left\{ \begin{smallmatrix} data,\ landmarks, \\ template,\ coupling \end{smallmatrix} \right\}} \mu_x \left(\sum_{k \in \{i,j\}} V_k^x(l_k) + \sum_{k,r \in \{i,j\}} V_{k,r}^x(l_k, l_r) \right)$$

2.1 Iconic Graph-Based Curve Tracking

The iconic part of our framework extends the MRF-based curve tracking method proposed in [6]. We represent the curve with a cubic B-spline based on points $\mathbf{c}(i;t)$ and we build an MRF graph containing one node per control point. We discretize the space of admissible displacements of each $\mathbf{c}(i;t)$ and denote with $\mathbf{v}(l_i)$ the admissible displacement associated with label l_i chosen for $\mathbf{c}(i;t)$. [Fig. (2)] illustrates this procedure. Like [6], let us note with $N_i(.)$ the basis function of each control point, $s \in [0,1]$ the curvilinear abscissa along the spline and $N_{i,i+1}(.)$ the following influence function: $N_{i,i+1}(s) = \frac{N_i(s)N_{i+1}(s)}{\sum_k N_k(s)N_{k+1}(s)}$. The spline curve $\mathcal{C}(s)$ is given by: $\mathcal{C}(s) = \sum_i N_i(s)\mathbf{c}(i;t)$. Let us denote with $C(s, l_i, l_{i+1})$ the curve obtained when the control points $\mathbf{c}(i;t)$ and $\mathbf{c}(i+1;t)$ are displaced by $\mathbf{v}(l_i)$ and $\mathbf{v}(l_{i+1})$ respectively (and similarly $C'(s, l_i, l_{i+1})$ its derivative), let us denote with $< .,. >$ the standard scalar product. In order to introduce the image support term, we consider the following exponential function: $\psi(x) = e^{-\gamma x}$.

We adopt two sums of pairwise potentials. The first forces the curve towards pixels likely to be part of the structure (the strength of the image support is important and the tangent of the curve is coherent with the local image orientation [1]),with $s(\mathbf{c}(i;t))$ being the curvilinear abscissa of $\mathbf{c}(i;t)$:

$$V_{i,i+1}^{data}(l_i, l_{i+1}) = \int_{s(\mathbf{c}(i;t))}^{s(\mathbf{c}(i+1;t))} \psi\left(\frac{| < C'(s, l_i, l_{i+1}), \mathcal{F}(C(s, l_i, l_{i+1}); t+1) > |}{||C'(s, l_i, l_{i+1})||} \right) ds$$

The second term penalizes the local changes of the curve derivatives with respect to a template $\mathcal{T}(s)$ corresponding to the form of the GW in the previous frames and updated using exponential forgetting: $\mathcal{T}(s) \leftarrow \frac{memory-1}{memory}\mathcal{T}(s) + \frac{1}{memory}C(s)$. It is more constraining than length preserving priors [12,6].

$$V_{i,i+1}^{template}(l_i, l_{i+1}) = \int_{s(\mathbf{c}(i;t))}^{s(\mathbf{c}(i+1;t))} \frac{||\epsilon\ C'(s, l_i, l_{i+1}) - \mathcal{T}'(s)||^2}{||\mathcal{T}'(s)||^2} ds$$

However, such an approach might require too many labels to deal with large displacements, can be sensitive to missing visual support and does not provide temporal understanding of the GW deformation. We introduced landmark matching into our framework for tackling these concerns.

2.2 Landmarks Matching

The landmark matching part of our system adopts a fully connected pairwise MRF, which optimal labeling indicates the candidates chosen for the matching. In order to increase the robustness and to deal with mis-detections or erroneous detections, we introduce a label $(L+1)$ corresponding to absent correspondences. We encoded a penalization of the deformation of the landmark configuration into the pairwise potentials. More precisely, the unary potentials were given by:

$$V_j^{landmarks}(l_j) = \begin{cases} K & if\ l_j = L+1 \\ 0 & otherwise \end{cases}$$

we chose the following potentials to penalize landmarks configurations changes:

$$V_{j,k}^{landmarks}(l_j, l_k) = \begin{cases} 0 & if\ l_j = L+1\ or\ l_k = L+1 \\ \min\left(||\mathbf{u}(j,l_j,k,l_k) - \mathbf{u}(j,k)||^2, \Gamma\right) & otherwise \end{cases}$$
$$\mathbf{u}(j,l_j,k,l_k) = \mathbf{p}^L(j;l_j;t+1) - \mathbf{p}^L(k;l_k;t+1)$$
$$\mathbf{u}(j,k) = \mathbf{p}(j;t) - \mathbf{p}(k;t)$$

This term aims at imposing geometric consistency between the location of landmarks in successive frames by considering their relative positions. The graph-matching cost increases with the Euclidean norm of the difference between the vector $\mathbf{u}(j,k)$ describing the configuration in the last frame and an admissible configuration $\mathbf{u}(j,l_j,k,l_k)$. If distances and orientations of landmarks pairs are preserved, then the cost is low. [Fig. (2)] presents example of matchings.

2.3 Coupled Markov Random Field for Hybrid Tracking

The hybrid model that we propose combines the two previous parts. The coupling term aims at imposing consistency between the iconic tracking and the geometric matching. As opposed to [13] where an approximation was used, we adopt an exact formulation that expresses consistency using singleton and pair-wise terms.

Let us denote with $\mathbf{w}_j(l_j)$ the displacement of the interest points $\mathbf{p}(j;t)$ corresponding to its matching with the candidate $\mathbf{p}^L(j;l_j;t+1)$ and with $s_j \in [0,1]$ its curvilinear abscissa. As in the previous section, $\mathbf{v}_i(l_i)$ denotes the displacement of the control point $\mathbf{c}(i;t)$ with respect to the frame t, and $N_i(.)$ the basis function of $\mathbf{c}(i;t)$. The following constraint is to be satisfied: $\sum_i N_i(s_j)\mathbf{v}_i(l_i) = \mathbf{w}_j(l_j)$ which imposes that the landmark displacement produced by the control points

motion is the same as the one determined from the matching. This constraint is converted into an energy term through penalization of the Euclidean distance:

$$E_j(\{l_i\}, l_j) = ||\sum_i N_i(s_j)\mathbf{v}_i(l_i) - \mathbf{w}_j(l_j)||^2 = \sum_{i \neq k} N_i(s_j)N_k(s_j) < \mathbf{v}_i(l_i), \mathbf{v}_k(l_k) > - 2\sum_i N_i(s_j) < \mathbf{v}_i(l_i), \mathbf{w}_j(l_j) > + \sum_i N_i(s_j)^2||\mathbf{v}_i(l_i)||^2 + ||\mathbf{w}_j(l_j)||^2$$

that converts the coupling constraint exactly into a sum of unary and pairwise potentials (contrary to [13] that minimizes only an upper bound of this energy).

Such an approach does not insure that the spline tracking will be guided by the geometric matching, since different displacements of the $\mathbf{c}(i;t)$ could lead to the same displacement of $\mathbf{p}(j;t)$. This can be easily addressed by adding pair-wise "rigidity" constraints to each landmark vicinity $v(j)$:

$$E_{r_j}(\{l_i\}) = \sum_{k \in v(j)} ||\mathbf{v}_k(l_k) - \mathbf{v}_{k+1}(l_{k+1})||^2$$

leading to the following coupling term:

$$\sum_k V_k^{coupling}(l_k) + \sum_{k,r} V_{k,r}^{coupling}(l_k, l_r) = \sum_j E_j(\{l_i\}, l_j) + \delta \sum_j E_{r_j}(\{l_i\})$$

3 Experimental Validation

In this section, we apply our approach to GW tracking in fluoroscopic sequences.

3.1 Steerable Filters Regularized by Tensor-Voting

We extracted the GW support with the most sensitive second order filter [9], which response is obtained (like Hessian response) by the eigen-decomposition of a matrix built using second order derivatives of the image. As we are interested in detecting dark structures, we considered the opposite of the main eigenvalue when it was negative (and a null response otherwise).

Towards improving their overall response, we used a fast variant of tensor voting (TV) introduced in [3]. Let us denote with $\hat{g}(\mathbf{x};t)$ and $\hat{\theta}(\mathbf{x};t)$ the response and the orientation provided by the steerable filter at location $\mathbf{x}$ in the frame t. We chose the following voting field (with notations [3], and we set $\sigma_{TV} = 4.5$):

$$V(r,\phi) = \frac{1}{16} e^{-\frac{r^2}{2\sigma_{TV}{}^2}} cos^4(\phi) \begin{pmatrix} 1 + cos(4\phi) & sin(4\phi) \\ sin(4\phi) & 1 - cos(4\phi) \end{pmatrix}$$

Like [3], we used this field to compute the new responses $g(\mathbf{x};t)$ and orientations $\theta(\mathbf{x};t)$, combined into the regularized field $\mathcal{F}(\mathbf{x};t)$:

$$\mathcal{F}(\mathbf{x};t) = (\mathcal{F}(\mathbf{x};t)_x, \mathcal{F}(\mathbf{x};t)_y) = (g(\mathbf{x};t)cos(\theta(\mathbf{x};t)), g(\mathbf{x};t)sin(\theta(\mathbf{x};t)))$$

[Fig. (1)] illustrates that the TV regularization clearly improves the results (contrary to the replacement of the Hessian by a steerable filter).

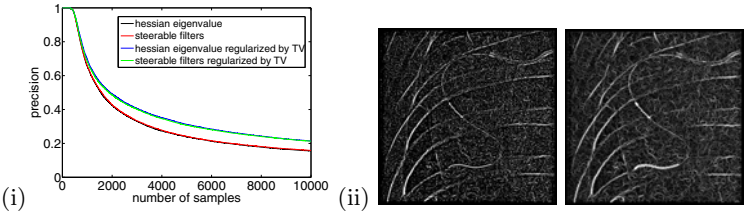


Fig. 1. (i) proportion of pixels extracted by Hessian and steerable filters regularized by TV belonging to the wire (estimated with 3 images by ranking pixels by decreasing filter response). (ii) norm of the steerable filter response and of the TV output.

3.2 Landmarks Extraction

Working with the regularized responses more reliable, we used the main eigenvalue of this tensor (inspired from structure tensor [5]) for landmarks detection:

$$T(\mathbf{x};t) = \sum_{\mathbf{y}} e^{-\frac{||\mathbf{x}-\mathbf{y}||^2}{2{\sigma_L}^2}} \begin{pmatrix} \mathcal{F}(\mathbf{y};t)_x^2 & \mathcal{F}(\mathbf{y};t)_x \mathcal{F}(\mathbf{y};t)_y \\ \mathcal{F}(\mathbf{y};t)_x \mathcal{F}(\mathbf{y};t)_y & \mathcal{F}(\mathbf{y};t)_y^2 \end{pmatrix}$$

This tensor favors pixels where strong and parallel responses accumulate. We set $\sigma_L = 2.5$. In order to take the local orientation into account when choosing the matching candidates $\mathbf{p}^L(j;m;t+1)$, we selected the L candidates $\mathbf{y}$ with the best matching scores (where $< .|. >$ denotes the inner product between matrices):

$$\mathscr{M}_j(\mathbf{y}) = < T(\mathbf{p}(j;t);t)|T(\mathbf{y};t+1) >$$

3.3 Implementation Details

First, we used a conventional multi-resolution strategy on the GW deformation search as suggested in [14]. We provided the spline at time 0 and the border of the field of view to our system and we forced the first control point to lie on this border if necessary. Besides, we updated the control points for keeping them equally distributed along the spline before processing every new frame in order to prevent a slow degeneracy of the spline. We made landmark detection more homogeneous along the GW by defining intervals of fixed length along the curve and choosing at most one landmark in each of them. FastPD [10] was used for minimizing E_{MRF}. We adressed abrupt elongations of the GW by elongating the tip of the curve in the direction of its tangent $\mathbf{d}$ before the control-points update. We appended pixels $\mathbf{q}$ while: $\psi(< \mathbf{d}, \mathcal{F}(\mathbf{q}) >) < 0.4$.

3.4 Validation

Validation was performed on 20 long fluoroscopic sequences of 200 frames of sizes between 512 and 1000 pixels covering a broad variety of clinical situations. The parameters were set using the first three sequences and performances were

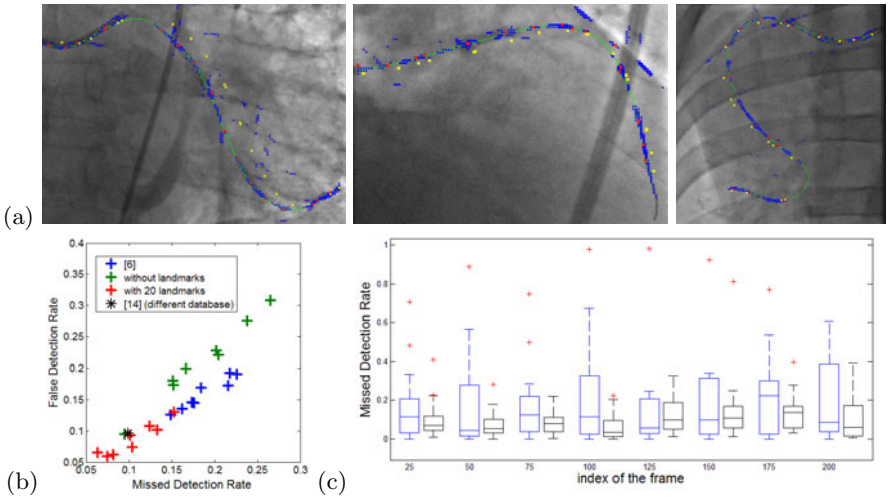


Fig. 2. (a) spline (green) landmarks detected along the previous spline (yellow) matching candidates (blue) and candidates chosen (red). (b) Mean tracking performances (measured every 25 frames for each experiment). [6] has been applied with a vesselness at scale 1.8 like our filters and $\lambda = 0.7$. The database for result [14] is different. (c) comparison of MDR obtained with [6] (blue boxes) and with our method (black boxes).

estimated on the other ones. We set $\mu_{data} = \mu_{template} = 1.0$, $\mu_{landmarks} = 2.10^{-4}$, $\mu_{coupling} = 3.10^{-5}$, $\delta = 0.1$, $\epsilon = 0.9$, $K = 45000$ for the landmarks not matched, displacements steps equal to $(25.0)2^{-\tau}$, $\tau \in \{0, \ldots 4\}$ and we sampled 25 displacements according to the sparse pattern [6]. We set $L = 24$, $\gamma = 6.10^{-3}$,$\Gamma = 500$ and $memory = 7$.

We measured missed detection rate (MDR) and false detection rate (FDR) for a distance of 5 pixels and for every 25 frames (2.5 seconds). These rates [14] correspond respectively to the proportion of pixels of the ground truth that lie too far from the spline and to the proportion of the spline that lie too far from the ground truth. [Fig. (2)] presents our results. [6] performs better than the iconic part of our model alone due to our worse displacement sampling and a different choice of initialization. Our complete method, however, outperforms [6] because the landmark matching both prevents the tracker from being mislead by local energy minima and preserves global spline configuration. We measured a distance between the tip of our tracker and the tip of the GW equal to 12.5 ± 11.5 pixels at the end of the sequences, what is acceptable since the GW tip is easy to locate. Our implementation processes one frame in 2 secondes (Intel Xeon 2.8 GHz) but would dramatically benefit from an adaptation to GPU.

4 Conclusion

In this paper we have presented a unified framework for geometric-iconic parametrized curves tracking. This novel framework achieves promising performances

on an extremely challenging task: the tracking of guide-wires in fluoroscopic sequences during cardiac angioplasty. One might investigate a GPU implementation achieving real-time and application of the method to a clinical setting. From a theoretical point of view, estimating uncertainties of the obtained solution might be helpful both for the quantitative interpretation of the result and for efficient parameterization of the search space. Another promising direction is to introduce long term memory through a dynamical system that could separate the different motion models.

References

1. Baert, S.A.M., Viergever, M.A., Niessen, W.J.: Guide-wire tracking during endovascular interventions. IEEE TMI 22(8), 965–972 (2003)
2. 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)
3. Franken, E., van Almsick, M., Rongen, P., Florack, L.M.J., ter Haar Romeny, B.M.: An efficient method for tensor voting using steerable filters. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3954, pp. 228–240. Springer, Heidelberg (2006)
4. Guy, G., Medioni, G.: Inferring global perceptual contours from local features. International Journal of Computer Vision 20, 113–133 (1996)
5. Harris, C.G., Stevens, M.J.: A combined corner and edge detector. In: 4th Alvey Vision Conference (1988)
6. Hauke Heibel, T., Glocker, B., Groher, M., Paragios, N., Komodakis, N., Navab, N.: Discrete tracking of parametrized curves. In: CVPR, pp. 1754–1761 (2009)
7. 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. LNCS, vol. 6363, pp. 440–448. Springer, Heidelberg (2010)
8. Isard, M., Blake, A.: Condensation conditional density propagation for visual tracking. IJCV 29 (1998)
9. Jacob, M., Unser, M.: Design of steerable filters for feature detection using canny-like criteria. IEEE PAMI 26(8), 1007–1019 (2004)
10. 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. CVIU 112, 14–29 (2008)
11. Pauly, O., Heibel, H., Navab, N.: A machine learning approach for deformable guide-wire tracking in fluoroscopic sequences. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6363, pp. 343–350. Springer, Heidelberg (2010)
12. Slabaugh, G., Kong, K., Unal, G., Fang, T.: Variational guidewire tracking using phase congruency. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part II. LNCS, vol. 4792, pp. 612–619. Springer, Heidelberg (2007)
13. 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. LNCS, vol. 6362, pp. 676–683. Springer, Heidelberg (2010)
14. Wang, P., Chen, T., Zhu, Y., Zhang, W., Zhou, S.K., Comaniciu, D.: Robust guidewire tracking in fluoroscopy. In: CVPR (2009)

Deformable Registration of Bronchoscopic Video Sequences to CT Volumes with Guaranteed Smooth Output

Tobias Reichl[1], Xiongbiao Luo[2], Manuela Menzel[3], Hubert Hautmann[3], Kensaku Mori[4], and Nassir Navab[1]

[1] Computer-Aided Medical Procedures, TUM, München, Germany
[2] Graduate School of Information Science, Nagoya University, Japan
[3] Medizinische Klinik I, Klinikum rechts der Isar, TUM, München, Germany
[4] Information and Communications Headquarters, Nagoya University, Japan

Abstract. We present a novel approach to tracking of flexible bronchoscopes by modeling the output as spatially continuous over time. Bronchoscopy is a widespread clinical procedure for diagnosis and treatment of lung diseases and navigation systems are highly needed. Tracking of the bronchoscope can be regarded as a deformable registration problem. In our approach we use hybrid image-based and electromagnetic tracking, and the bronchoscope pose relative to CT data is interpolated using Catmull-Rom splines for position and SLERP for orientation.

We evaluate the method using ground truth poses manually selected by experts, where mean inter-expert agreement was determined as 1.26 mm. For four dynamic phantom data sets, the accuracy of our method is between 4.13 and 5.93 mm and shown to be equivalent to previous methods. We significantly improve inter-frame smoothness from 2.35–3.08 mm to 1.08–1.51 mm. Our method provides a more realistic and physically plausible solution with significantly less jitter. This quantitative result is confirmed by video output, which is much more consistent and robust, with fewer occasions of tracking loss or unexpected movement.

1 Introduction

Flexible bronchoscopy is a widespread clinical procedure for diagnosis and treatment of lung diseases, and one of its most common applications is transbronchial biopsy. This is commonly performed after lesions have been identified on Computed Tomography (CT) images, so it will definitely be beneficial to transfer this 3D information to the operating room. Since bronchoscopy is an inherently monitor-based procedure, augmentation of video images with guidance or targeting information is straightforward and promises high clinical acceptance due to smooth integration into the clinical workflow. The combination of flexible bronchoscopy with electromagnetic tracking (EMT) was first reported by Solomon et al. [1], and hybrid image-based and EM tracking was proposed by Mori et al. [2], and was improved by Luo et al. [3,4] and Soper et al. [5]. Hybrid tracking

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 17–24, 2011.

complements the advantages of image-based tracking and EMT, since image-based tracking is inaccurate for frames with little to no information due to low contrast. EMT suffers from noise and artifacts while the sensor is moved [6], and relative errors when anatomy is moving due to respiration or heartbeat.

While accuracy and computational speed are clearly important for real-time applications like intraoperative navigation, smoothness of the output is no less relevant. Less jitter of augmented reality overlays increases operator comfort and acceptance. Smooth output may provide benefits when further processing the output like for motion models, or temporal and spatial synchronization of multiple video sequences in the case of repeat examinations or clinical studies.

Tracking of the bronchoscope can be seen as a deformable registration problem as shown below, and in this work we introduce a novel solution to this problem. The deformable registration of the images over the bronchoscope trajectory is performed using Catmull-Rom splines and spherical linear interpolation (SLERP). Smoothness between consecutive frames is ensured using interpolation. We present a thorough and quantitative evaluation with respect to expert-provided ground truth data, including the determination of intra- and inter-expert agreement, and a quantitative comparison to the state of art.

2 Methods

A continuous description of bronchoscope pose at time t is given by its position $p(t)$ and orientation $q(t)$. Since both the real movement of the bronchoscope and the movement of anatomy (breathing, heartbeat, etc.) are spatially smooth over time, movement of the bronchoscope relative to patient anatomy is smooth as well.

Discrete control points for the trajectory are set equidistant over time with spacing s. For interpolation of position we use Catmull-Rom splines [7], because the resulting curve is continuously differentiable and passes directly through the control points:

$$p(t) = \frac{1}{2}\begin{pmatrix}1 & u & u^2 & u^3\end{pmatrix}\begin{pmatrix}0 & 2 & 0 & 0\\ -1 & 0 & 1 & 0\\ 2 & -5 & 4 & -1\\ -1 & 3 & -3 & 1\end{pmatrix}\begin{pmatrix}p_{i-1}\\ p_i\\ p_{i+1}\\ p_{i+2}\end{pmatrix}, \tag{1}$$

where $p_{i-1} \dots p_{i+2}$ are positions of consecutive control points, $i = \lfloor t/s \rfloor$ is a control point index, and $u = t/s - \lfloor t/s \rfloor$ is the interpolation ratio between control points p_i and p_{i+1}.

For orientation we use quaternions, because they allow a continuous representation. Then, for interpolation between quaternions we use SLERP [8]:

$$q(t) = q_i(q_i^{-1}q_{i-1})^u = \frac{\sin(1-u)\theta}{\sin\theta}q_i + \frac{\sin u\theta}{\sin\theta}q_{i+1}, \tag{2}$$

where θ is the rotation difference between q_i and q_{i-1}. Initial parameters p_k and q_k for all control points are taken directly from the EMT measurements.

For the matching between bronchoscope images and CT image, we employ as an intermediate step virtual bronchoscopic images $I_V(p_k, q_k)$, where the CT image is rendered for a camera with position p_k and orientation q_k. We seek to maximize

$$E(p,q) = \underbrace{S(p,q)}_{\text{similarity}} - \underbrace{\lambda_1 \cdot R_T(p,q)}_{\text{tension}} - \underbrace{\lambda_2 \cdot R_B(p)}_{\text{bending}}, \tag{3}$$

where λ_1 and λ_2 are weighting parameters. The similarity $S(p,q)$ between real and virtual images is computed via the Local Normalized Cross Correlation (LNCC)

$$S(p,q) = \sum_k \text{LNCC}(I_R(t_k), I_V(p_k, q_k)), \tag{4}$$

where k is an image index, $I_R(t_k)$ is the real image at time t_k, and $I_V(p_k, q_k)$ is the virtual bronchoscopic image. The patch size for LNCC was set to 11. The gradient of $S(p,q)$ for movement for each single control point i can be approximated from the frames in its support $(i-2, i+2)$ via finite differences. Frames outside its support are not influenced by movement of this control point, so gradient computation can be decoupled for each control point. In addition, since the total number of frames is constant, the total computational effort does not depend on the number of control points.

According to Hooke's law, spring force is proportional to displacement, so we model tension, the cost of distance from the EM tracking measurements, as

$$R_T(p,q) = \sum_k \|p_k - p_{k,0}\| + \alpha \cdot \theta(q_k, q_{k,0}), \tag{5}$$

where $p_{k,0}$ and $q_{k,0}$ are the position and orientation measured by EM tracking at time t_k, and the rotation difference $\theta(\cdot)$ is computed as the angular component of the difference quaternion. α is the ratio between rotational and translational spring constants and was set to $3.10°/mm$, since this was the ratio of errors observed with the human experts when recording ground truth data.

In addition to the inherent smoothness of the spline curve, another term penalizes large translations between control points. According to Euler-Bernoulli beam theory, the bending moment of the trajectory is proportional to its curvature, so we choose analogously

$$R_B(p) = \sum_k \left\| \nabla_k^2 p_k \right\|. \tag{6}$$

We iteratively estimate the optimal parameters for $E(p,q)$ by gradient descent. The update p_u, q_u is given by

$$\nabla E(p,q) = \nabla S(p,q) - \lambda_1 \cdot \nabla R_T(p,q) - \lambda_2 \cdot \nabla R_B(p,q) \tag{7}$$

$$(p_u, q_u) \leftarrow sign(\nabla E(p,q)) \cdot min(\tau \cdot \|\nabla E(p,q)\|, \delta), \tag{8}$$

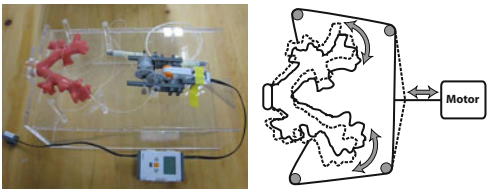


Fig. 1. Dynamic motion phantom (left) and description of phantom operation (right)

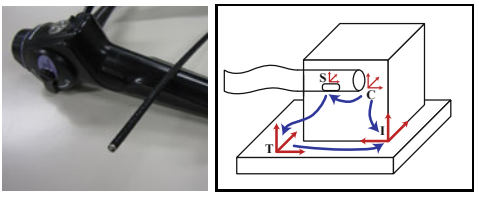


Fig. 2. Bronchoscope with embedded EMT sensor (left) and coordinate systems involved (right): bronchoscope camera (C), CT image (I), electromagnetic tracking (T), and tracking sensor (S) coordinate frames. The transformation ${}^{T}T_{S}$ is measured, the transformations ${}^{S}T_{C}$ and ${}^{I}T_{T}$ are calibrated, and the transformation ${}^{I}T_{C}$ is optimized for each frame

where τ is a magnitude control parameter, and δ is the maximum step width. During optimization, only data for a neighborhood of frames needs to be available, so frames can be processed in sequence like with previous approaches.

3 Evaluation and Results

Setup: We use a 3D Guidance EMT system (Ascension, Burlington, USA) and a BF-P260F flexible fiberoptic bronchoscope (Olympus, Tokyo, Japan). One EMT sensor was fixed inside the bronchoscope working channel.

The dynamic phantom is a CLA 9 (CLA, Coburg, Germany), which was chosen due to its closely human-mimicking surface. It was connected to a motor (Lego, Billund, Denmark) via nylon threads (cf. Fig. 1). Four data sets consisting of video sequences and EMT data recordings were acquired with different amplitudes of simulated breathing motion between 7.48 and 23.65 mm.

A CT scan of the phantom was acquired with 0.5 mm slice thickness. The bronchoscope and the different coordinate systems and transformations are shown in Fig. 2. For point-based registration between CT and EMT coordinate systems, 29 external landmarks were used, average residual error was 0.92 mm. Camera intrinsics, deformation, and hand-eye calibration were performed using established methods [9].

Both video sequence and CT data were stored in graphics processing unit (GPU) memory, and virtual bronchoscopic image rendering as well as similarity were

Table 1. Accuracy comparison of approaches. In each cell, mean error and standard deviation with respect to ground truth are given for translation as well as rotation. Expert agreement denotes inter-expert standard deviation for each particular dataset. The results are equivalent with regard to expert agreement.

Accuracy	Data set 1 n=116	Data set 2 n = 113	Data set 3 n = 110	Data set 4 n = 107
Solomon et al. [1]	4.90 ± 2.41 mm $8.99 \pm 5.78°$	5.86 ± 2.53 mm $9.50 \pm 4.13°$	5.71 ± 2.53 mm $10.60 \pm 4.59°$	7.00 ± 3.22 mm $13.09 \pm 10.16°$
Mori et al. [2]	4.57 ± 2.57 mm $9.13 \pm 5.86°$	5.76 ± 2.65 mm $9.52 \pm 4.34°$	5.34 ± 2.95 mm $10.32 \pm 4.86°$	6.69 ± 3.46 mm $13.24 \pm 10.43°$
Luo et al. [3]	4.12 ± 2.65 mm $9.29 \pm 5.69°$	5.11 ± 2.76 mm $9.10 \pm 4.48°$	4.77 ± 2.98 mm $10.29 \pm 5.18°$	6.68 ± 4.78 mm $14.33 \pm 11.82°$
Luo et al. [4]	3.95 ± 2.78 mm $9.13 \pm 5.76°$	4.00 ± 2.65 mm $9.58 \pm 4.17°$	4.80 ± 3.28 mm $10.65 \pm 4.50°$	5.32 ± 3.58 mm $13.13 \pm 10.13°$
Proposed Method	4.13 ± 2.30 mm $9.79 \pm 4.99°$	4.65 ± 2.58 mm $10.81 \pm 5.22°$	4.93 ± 2.41 mm $11.08 \pm 4.67°$	5.93 ± 2.97 mm $14.26 \pm 9.49°$
Expert agreement	1.27 mm 5.21°	1.15 mm 3.64°	1.33 mm 4.63°	1.27 mm 5.23°

computed on GPU using OpenGL. Experiments were conducted on a standard workstation (Intel Core Duo 6600, NVidia GeForce 8800 GTS).

As an extension of the "manual registration" approach by Soper et al. [5] we chose an evaluation based on expert-provided ground truth data, since then a direct deduction of clinical relevance is possible, and from multiple experts the limits of image-based methods might be learned.

Ground truth data was independently and repeatedly collected by two experts, one expert bronchoscopist (A) and one scientist (B). For each real image, the position and orientation of the virtual bronchoscopic image were manually adapted via mouse and keyboard, until they matched the real image as closely as possible. Each expert was blinded to the other expert's results, as well as to his own results from previous sessions. Due to the time-consuming process, only approximately every eighth frame was matched. First, pose data from multiple sessions was averaged per expert, then pose data from both experts.

Intra-expert agreement (mean standard deviation) was 1.66 mm and 5.80° (A) and 1.44 mm and 3.94° (B). Inter-expert agreement was 1.26 mm and 4.78°. The ratio between intra- and inter-expert agreement indicates considerable overlap between the experts' results. These margins might indicate a limit for any approach based on registration of real and virtual bronchoscopy images.

We compare the proposed method to our own implementations of four previously published approaches, which have already been applied to similar trajectories: bronchoscope tracking by EMT only [1], intensity-based registration (IBR) with direct initialization from EMT [2], IBR with dynamic initialization from EMT [3], and IBR with a Sequential Monte Carlo sampler based on EMT [4]. Quantitative results for the accuracy are given in Table 1. The precision of our evaluation method is defined by expert agreement, and within those bounds the accuracy of the proposed method is equivalent to previous approaches.

Table 2. Smoothness comparison of approaches. In each cell, mean inter-frame distance and standard deviation are given for translation as well as rotation. Smoothness is significantly improved by the proposed method.

Smoothness	Data set 1	Data set 2	Data set 3	Data set 4
Solomon et al. [1]	2.86 ± 1.62 mm $2.11 \pm 8.21°$	3.47 ± 1.72 mm $3.84 \pm 8.59°$	3.01 ± 1.65 mm $3.81 \pm 11.25°$	3.46 ± 1.73 mm $3.83 \pm 8.81°$
Mori et al. [2]	3.35 ± 2.09 mm $3.50 \pm 9.68°$	3.87 ± 2.31 mm $4.74 \pm 8.63°$	3.61 ± 2.18 mm $6.05 \pm 14.51°$	4.04 ± 2.35 mm $5.63 \pm 10.74°$
Luo et al. [3]	2.86 ± 2.11 mm $3.22 \pm 10.42°$	3.43 ± 2.30 mm $4.54 \pm 8.83°$	3.11 ± 2.20 mm $5.17 \pm 12.35°$	3.54 ± 2.35 mm $5.36 \pm 10.69°$
Luo et al. [4]	2.35 ± 1.88 mm $2.12 \pm 8.24°$	2.95 ± 2.09 mm $3.94 \pm 8.92°$	2.71 ± 2.08 mm $3.84 \pm 11.32°$	3.08 ± 2.28 mm $3.96 \pm 9.35°$
Proposed Method	1.23 ± 0.76 mm $1.96 \pm 5.82°$	1.51 ± 1.04 mm $3.18 \pm 6.23°$	1.12 ± 0.69 mm $3.22 \pm 11.71°$	1.08 ± 0.77 mm $3.62 \pm 9.66°$

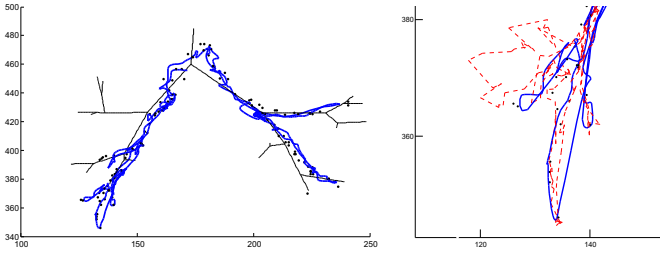


Fig. 3. Left: frontal view of airway structure, trajectory from proposed method (solid blue), and ground truth positions (black dots). Right: close-up view of trajectories from previous approach [4] (dashed red) and proposed method (solid blue), showing much smoother output of the latter

In Table 2 quantitative results for the smoothness of the output are given in terms of inter-frame distances. The results for previous methods agree with Soper et al. [5], who reported mean distances of 2.37 mm and 8.46°. The proposed method shows significantly better smoothness. In order to also visualize the difference, two output trajectories are shown in Fig. 3. In particular, the close-up view shows significantly less jitter in the output of the proposed method than of the previous approach.

Computation time is 0.98 seconds per frame, i.e. between 12.8 and 14.9 minutes for the 781 to 911 frames of our video sequences. For other methods a run time between 0.75 and 2.5 seconds [2,3,5,4] per frame is reported, so our computation speed is equivalent to previous approaches.

4 Discussion

The accuracy of our method is equivalent to previous approaches, with a significant improvement in smoothness. The actual limit to smoothness is unknown,

but we attribute most of the remaining inter-frame distance to the real bronchoscope motion. Due to the smoothness constraint, our method by design provides a more realistic and physically plausible solution to the tracking problem, which has significantly less jitter. This quantitative result is confirmed by video output[1], which is much more consistent and robust, with fewer occasions of tracking loss or unexpected movement. The filtering effect removes noise from the input, but can also lead to some loss of detail, if e.g. the direction of motion changes.

Soper et al. [5] report a slightly lower error of 3.33 mm for their method of tracking an ultrathin bronchoscope. However, the actual clinical relevance of ultrathin bronchoscopes is low, as those have not yet proven diagnostic advantages over conventional bronchoscopes [10,11] and do not provide any steering mechanism. Also, Soper et al. only considered comparatively simple trajectories consisting of one smooth, one-way motion from the trachea into the peripheral lung. In contrast, recommendations for bronchoscopic procedures commonly list e.g. the need to visually inspect all accessible branches at the beginning of an intervention [12]. A typical trajectory will therefore consist of much more irregular motion and more changes of direction.

We expect our method to be robust against transient artifacts occurring in real clinical images, like specular reflections or bubbles, since groups of frames are matched and smoothness is enforced.

In the current phantom design movement is mostly within the coronal plane and left-right. However, our method does not impose any assumptions on the breathing motion, which, depending on the degree of patient sedation, can be irregular and interrupted by coughing or choking.

Future work will definitely include acquisition of ground truth data from more interventional experts. However, in clinical routine time is a precious resource and the acquisition process is extremely time-consuming. When working accurately, approximately 30-40 poses can be recorded per hour. As an alternative ground truth, landmarks could have been determined in the CT image and videos, but then only in a fraction of the video images those landmarks are visible, and especially the performance of a method in the absence of landmarks is relevant.

The next step will be applying the proposed method to trajectories recorded during actual interventions in humans. Since the dynamic motion phantom sequences were recorded anticipating the real procedure, data acquisition can be brought to the operating room with minimal interruption to the surgical workflow.

5 Conclusions

We present a novel approach to hybrid imaged-based and electromagnetic tracking of flexible bronchoscopes by modeling the output as spatially smooth over time. While providing equivalent accuracy at equivalent speeds, we significantly improve smoothness. As we only impose minimal prior knowledge about the

[1] http://campar.in.tum.de/files/publications/reichl2011miccai.video.avi

visual appearance of anatomy, but do not depend on e.g. airway segmentation, tree structure, or bronchoscope properties, our method can be applied to other endoscopic applications as well.

Acknowledgements. This work was supported by Deutsche Forschungsgemeinschaft under grants NA 620/2-1 and 446 JAP 113/348/0-1, European Union FP7 under grant n°256984, JSPS Grant for Scientific Research, and the TUM Graduate School of Information Science in Health (GSISH). Thanks to M. Feuerstein for help with Fig. 3.

References

1. Solomon, S.B., White Jr., P., Wiener, C.M., Orens, J.B., Wang, K.P.: Three-dimensional CT-guided bronchoscopy with a real-time electromagnetic position sensor: a comparison of two image registration methods. Chest 118, 1783–1787 (2000)
2. Mori, K., Deguchi, D., Akiyama, K., Kitasaka, T., Maurer Jr., C.R., Suenaga, Y., Takabatake, H., Mori, M., Natori, H.: Hybrid bronchoscope tracking using a magnetic tracking sensor and image registration. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3750, pp. 543–550. Springer, Heidelberg (2005)
3. Luo, X., Feuerstein, M., Sugiura, T., Kitasaka, T., Imaizumi, K., Hasegawa, Y., Mori, K.: Towards hybrid bronchoscope tracking under respiratory motion: evaluation on a dynamic motion phantom. In: Proc. SPIE, vol. 7625, p. 76251B (2010)
4. Luó, X., Reichl, T., Feuerstein, M., Kitasaka, T., Mori, K.: Modified hybrid bronchoscope tracking based on sequential monte carlo sampler: Dynamic phantom validation. In: Kimmel, R., Klette, R., Sugimoto, A. (eds.) ACCV 2010, Part III. LNCS, vol. 6494, pp. 409–421. Springer, Heidelberg (2011)
5. Soper, T., Haynor, D., Glenny, R., Seibel, E.: In vivo validation of a hybrid tracking system for navigation of an ultrathin bronchoscope within peripheral airways. IEEE Transactions on Biomedical Engineering 57, 736–745 (2010)
6. Hummel, J., Figl, M., Bax, M., Shahidi, R., Bergmann, H., Birkfellner, W.: Evaluation of dynamic electromagnetic tracking deviation. In: Proc. SPIE, vol. 7261, p. 72612U (2009)
7. Catmull, E., Rom, R.: A class of interpolating splines. Computer Aided Geometric Design, 317–326 (1974)
8. Shoemake, K.: Animating rotation with quaternion curves. In: SIGGRAPH (1985)
9. Wengert, C., Reeff, M., Cattin, P., Szekely, G.: Fully automatic endoscope calibration for intraoperative use. In: Bildverarbeitung in der Medizin (2006)
10. Tachihara, M., Ishida, T., Kanazawa, K., Sugawara, A., Watanabe, K., Uekita, K., Moriya, H., Yamazaki, K., Asano, F., Munakata, M.: A virtual bronchoscopic navigation system under X-ray fluoroscopy for transbronchial diagnosis of small peripheral pulmonary lesions. Lung Cancer 57, 322–327 (2007)
11. Yamamoto, S., Ueno, K., Imamura, F., Matsuoka, H., Nagatomo, I., Omiya, Y., Yoshimura, M., Kusunoki, Y.: Usefulness of ultrathin bronchoscopy in diagnosis of lung cancer. Lung Cancer 46, 43–48 (2004)
12. Häussinger, K., Ballin, A., Becker, H.D., et al.: Recommendations for quality standards in bronchoscopy. Pneumologie 58, 344–356 (2004)

MR-Based Real Time Path Planning for Cardiac Operations with Transapical Access

Erol Yeniaras[1], Nikhil V. Navkar[1,2], Ahmet E. Sonmez[1], Dipan J. Shah[3], Zhigang Deng[2], and Nikolaos V. Tsekos[1]

[1] Medical Robotics Lab
[2] Computer Graphics and Interactive Media Lab,
Department of Computer Science, University of Houston, Houston, TX 77004, USA
[3] Methodist DeBakey Heart & Vascular Center, Houston, TX 77030, USA
{yeniaras,nvnavkar,aesonmez,ntsekos,zdeng}@cs.uh.edu,
djshah@tmhs.org

Abstract. Minimally invasive surgeries (MIS) have been perpetually evolving due to their potential high impact on improving patient management and overall cost effectiveness. Currently, MIS are further strengthened by the incorporation of magnetic resonance imaging (MRI) for amended visualization and high precision. Motivated by the fact that real-time MRI is emerging as a feasible modality especially for guiding interventions and surgeries in the beating heart; in this paper we introduce a real-time path planning algorithm for intracardiac procedures. Our approach creates a volumetric safety zone inside a beating heart and updates it on-the-fly using real-time MRI during the deployment of a robotic device. In order to prove the concept and assess the feasibility of the introduced method, a realistic operational scenario of transapical aortic valve replacement in a beating heart is chosen as the virtual case study.

Keywords: Real time MRI, Image Guided Surgeries, and Beating Heart.

1 Introduction

Contemporary improvements in the field of medical robotics, and a series of successful clinical applications, have led to the emergence of interventional robots by the clinical and technical community. The inclusion of real-time image guidance in robotic-assisted interventions may further elevate the field by offering improved information-rich visualization, as well as option of assessing the tissue before, during and after a procedure [1]. Considering the challenges associated with the continuous cardiac motion, real-time image guidance can provide a number of benefits especially for robot-assisted surgeries in a beating heart [2, 3].

Among the emerging clinical paradigms in the area of minimally invasive procedures in a beating heart is magnetic resonance imaging (MRI) guided prosthetic aortic valve implantation via transapical access; a procedure that has been demonstrated manually [4] and with robotic assistance [2]. For such off-pump procedures, an important factor is the efficacy of using images to assess the dynamic environment during operation. One effective method is the extraction of dynamic

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 25–32, 2011.

access corridors from images [5]. MRI was selected by numerous investigators for its high soft-tissue contrast, absence of ionizing radiation, and inherent robustness and versatility [4, 5]. With the current state-of-the-art MRI, dynamic images can be collected at a rate of 40-50 ms/image. Practically, only a single imaging plane can be collected with such a high, for MRI standards, rate. Thus the question is how a 3D corridor inside the beating heart can be updated using a single plane.

To address this issue, in this work, we evaluate a method that combines preoperative multislice dynamic MRI (i.e., cine MRI) with single-slice real-time MRI to update an access corridor from the apex to the aortic annulus. Cine MRI is used to generate a preoperative 3D corridor in the left ventricle (LV) which is updated on-the-fly by registering it onto the intraoperative real-time MR images. The method was assessed for accuracy of the corridor registration and simulated for the deployment of a virtual robot for 12 subjects.

2 Methodology

In a typical transapical scenario, the robotic manipulator enters LV via a trocar affixed at the apex, T(t); and maneuvers toward the targeted center of aortic annulus, A(t) as depicted in Fig. 1. Preliminary analyses of the cine data from 12 healthy volunteers indicated that LV can be transversed with a cylindrical corridor and the deployment path from T(t) to A(t) is not a straight line. For a precise orthogonal approach to aortic annulus, a dynamic bending point, B (t), near the base of LV is needed. The characteristics of the corridor, as well as the aortic annulus diameter, coronary ostial anatomy and apical entrance point were determined from cine MRI, whereas real-time MRI was used to update the corridor and follow the operation.

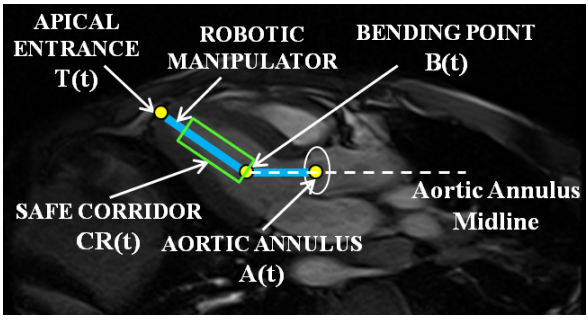


Fig. 1. A long axis MR image shows a typical transapical approach to a beating heart at time t

2.1 Preoperative Planning

For any given cardiac phase (i.e., at time frame t), the access corridor CR(t) is defined as the largest cylindrical volume that lies along LV from the apex toward the base of the heart. Then, an appropriate-sized surgical device should be able to deploy inside CR(t) from T(t) to A(t) without colliding or injuring the endocardium, papillary muscles or chordae tendinae. The cylindrical corridor was generated from cine

datasets (n=12) collected with a true fast imaging, steady-state precession pulse sequences (TrueFISP) with TR/TE = 2.3 ms/1.4 ms, flip angle = 80°, slice thickness = 6.0 mm, and acquisition matrix = 224x256. Each dataset included 19 short axes (SA) and 5 long axes (LA) slices, capturing heart motion with 25 frames over a complete cardiac cycle. In order to determine the transient positions of the corridor and the deployment points (i.e., T(t), B(t), and A(t)), SA and LA images were segmented using a region-growing algorithm based on Insight Toolkit (ITK) filters to extract the apex, LV and aortic annulus. As shown in Fig. 2(a) and 2(b), to realistically model the corridor, papillary muscles and chordae tendinae were also considered. For every single heart phase t (t=1 to 25), CR(t) is constructed as follows:

1. The SA slices with visible blood pool are determined by checking the inside surface areas of LV segmentations, i.e., selecting the non-zero ones as in Fig. 2(b);
2. LV segmentation contours of these SA slices are projected onto a virtual plane along their common orthogonal axis to find their intersection polygon (INP_T) by 2D polygon clipping. This projection is based on the fact that SA slices are parallel to each other and collected with the same field of view;
3. The largest circle (S_T) that fits into INP_T is determined. Since a circle can be created with a center point and a radius, let`s define C_T as the Center of S_T and, R_T as the Radius of S_T. Then the centroid of INP_T is chosen as the center of S_T, shown in Eq. (1) where N is the number of edges of INP_T:

$$C_T(x, y) = \frac{1}{N} \sum_{i=1}^{N} INP_T(x_i, y_i) \quad (1)$$

   Thus, R_T can be safely defined as "the minimum distance from C_T to the edges of INP_T" as formulated in Eq. (2):

$$R_T = \min_{1...N} (\| C_T(x, y) - INP_T(x_i, y_i) \|) \quad (2)$$

4. Finally, S_T is stretched from the apex to the base of the heart, through all SA slices to generate a circular straight access corridor as shown in Fig. 2(c).

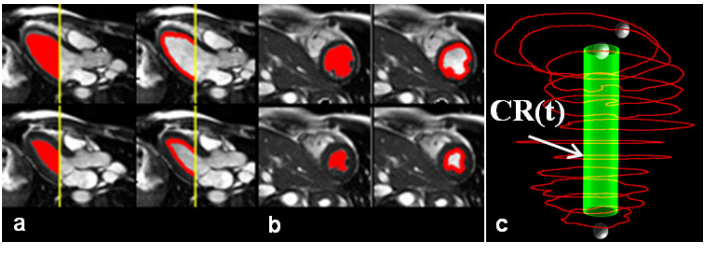


Fig. 2. Selected segmentations of LA (a), SA (b) slices from diastole and systole phases highlight LV blood pool and boundaries, and a sample corridor with endocardial contours (c)

Since a unique 3D access corridor is generated for every single heart phase, CR(t) is a 4D dynamic entity defined for a full heart cycle consisting of 25 time frames. The final step of preoperative planning is to set the initial positions of deployment points as follows: (1) T(t) is selected manually by a cardiovascular surgeon as the tip of the LV (the apex) on the central LA slice (and verified computationally that it belongs to

a SA slice that shows only myocardium with no LV blood pool); (2) A(t) is determined from the segmentation contours of two LA and one SA slices that include the aortic valve annulus at the level of aortic valve leaflets. Note that, the initial selection of the exact target point is also made by the surgeon (or cardiologist); (3) B(t) is assigned as the intersection of the aortic annulus midline and the top face of the safe corridor initially (as also depicted in Fig. 1).

2.2 Intraoperative Guidance

In this phase, the access corridor and deployment points, generated preoperatively for a complete heart cycle, are updated on-the-fly. Intraoperative guidance is based on continuous real-time acquisitions of a central LA slice taken from the very same healthy volunteers (n=12) but this time spanning 30 full heart-beats with breathing (TrueFISP parallel imaging with effective repetition time = 48.4 ms, TE=0.95 ms, alpha=65^{o}, slice thickness=6.0 mm, acquisition matrix = 160x66). After comprehensive analyses of MRI data and different imaging planes, we observed that heart mainly translates along and rotates around its long axis without significant out-of-plane motion with respect to the real-time LA slice under consideration. Moreover, breathing adds an extra vertical motility with respect to the MR table, as denoted with R in Fig. 3, which can also be followed on the same LA slice effectively. To the end, we choose this single real-time LA slice for intraoperative guidance.

The most challenging task is to register the preoperative corridor onto the real-time LA slice on-the-fly during the operation. This is done via two major steps: (1) Determine the heart phase in which the real-time slice was collected (and thus match it with the corresponding corridor); (2) Adjust the position and orientation of this corridor to account for heart motion due to respiration, arrhythmias, etc.

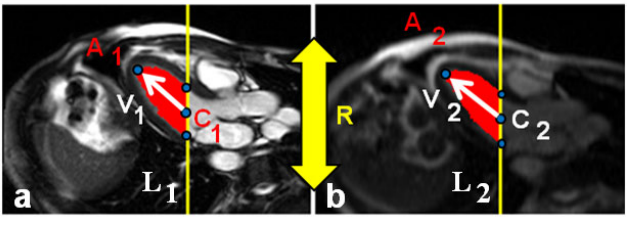


Fig. 3. Illustration of LA segmentations depicting the blood-pool area, R: respiratory motion; A_J: Apex points; C_J: Midpoints; L_J: Vertical lines, V_J: $C_J \rightarrow A_J$ LV directional vectors for non-real-time and real-time LA slices respectively; J = 1 and 2

First, LV is segmented in the real-time LA image using the same region-growing algorithm and area of the blood-pool is calculated. Then, this area is compared to each of the 25 preoperative LV areas of the same LA slice that is extracted with the same parameters, to find the closest. This comparison is done by a conservative approach, i.e., selecting the closest one with the minimum size to guarantee the aforementioned safety criterion. Once it is found, the corridor corresponding to this heart phase is selected as the one to be registered to the real-time LA slice. Fig. 3 shows a segmented cine (non real-time) LA slice and its real-time counterpart respectively.

Next, the corridor needs to be correctly positioned onto the real-time slice. To achieve that, a vertical line (L_2) crossing the base of the heart is defined. This line is defined such that segmentation includes most of LV blood pool but not the aortic valve. Then, the intersection points of this line to the endocardial wall are determined and their midpoint (C_2) is calculated as shown in Fig. 3. The same operation is performed for the corresponding preoperative image to compute L_1 and C_1, and thus we can compare C_1 with C_2. It should be noted that, A_1 and A_2 denote the apex points, while V_1 and V_2 are directional vectors for the LVs.

Then, the relative displacement between C_1 and C_2 is calculated and applied to the top-center point of the corridor, P_{11}, to find P_{21} as shown in Fig.4 (a). To adjust the orientation, the angle between V_1 and V_2 is calculated and applied to the direction vector of the preoperative corridor V_{1C} to find V_{2C}. Finally, the resultant corridor is registered to its new position as depicted in Fig.4 (b) and it is ready for the robotic manipulator. The above process is highly efficient and real-time (i.e., all the computing takes less than 48.4 ms, which is less then effective repetition time).

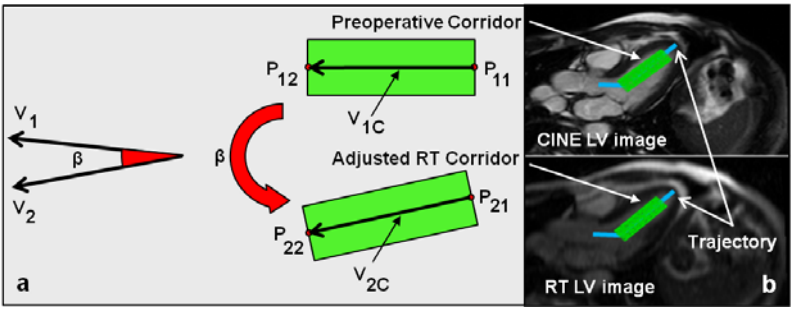


Fig. 4. (a) Corridor registration using vectors; and (b) the registered trajectory on LA slices

Finally, initial deployment points are tracked during the (virtual) operation using an efficient 2D fast-tissue-tracking algorithm [6]. In an *in vivo* scenario, the target and the bending points can be tracked with such an algorithm whereas appropriate MRI methods should provide the position of the apex point, T(t) in real-time (e.g., miniature RF coil beacons on the trocar [7]).

2.3 Experimental Studies

The proposed method was tested for the registration accuracy of the corridor onto the real-time MRI. Specifically, the registered corridors were compared to the ground-truths that were created by manually locating them onto their correct positions in all the real-time images for 30 full heart cycles (n=12). Fig. 5 shows three examples of corridor positioning errors.

We also simulated the deployment of a six degrees-of-freedom (DOF) RRPRRP (R: Rotational joint, P: Prismatic joint) virtual robot inside the registered corridor to assess the possible collisions with the endocardium or the aortic walls. The proximal end of the robot was assumed to be attached to T(t) with a continuous actuation to follow the motion of the apex and Denavit-Hartenberg convention was used in inverse

kinematics to deploy the robot along the corridor. The inputs for robot control were the dynamic coordinates of deployment points as well as the initial conditions specified by the operator, e.g., the time frame when the robot initiates its maneuvering, and whether and for how long it may hold a certain position along its path. Motion of the virtual robot entails the following three steps: deployment of the first link from the apex toward B(t), extension of the second link toward A(t), and the holding of the position. Since the device must stay inside CR(t), turn can occur before or after the initial B(t) provided that the first section remains in the corridor and the tip heads to the center of aortic annulus. During the maneuvering process, the robot control supplies the values of the updated DOFs for each time instance. Robot deployment was visually simulated along with the surgical field using OpenGL.

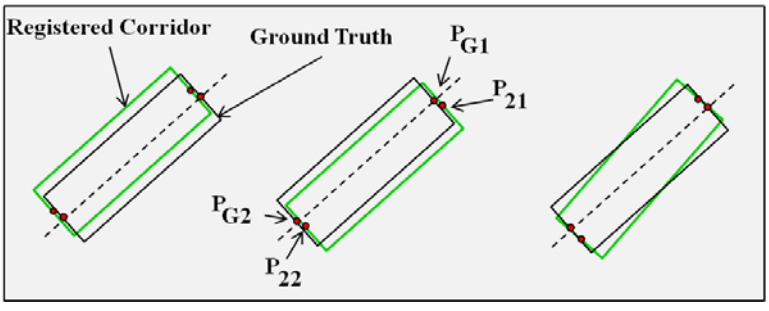


Fig. 5. Registered points: P_{21}, P_{22} and ground-truth: P_{G1}, P_{G2} for three different samples

3 Results and Discussion

Application of the method on all preoperative cine sets (n=12) demonstrated that a dynamic cylindrical corridor can be defined and tracked for safe deployment inside the LV of the beating heart. The average base diameter of this corridor for 12 subjects was 9 mm in systole and 22 mm in diastole. In regard to the code for on-the-fly processing of the real-time MRI, corridor selection was practically error-free as a reflection of the conservative approach in selecting them (i.e., the minimal size). The average distal error for the starting point of the corridor (P_{21}) was 1.3 mm while it was 2.0 mm for the ending point (P_{22}) as formulated in Eq. (3) and listed in Table 1.

$$\varepsilon_S = \| P_{21} - P_{G1} \| ; \varepsilon_E = \| P_{22} - P_{G2} \| \tag{3}$$

This difference is mainly caused by the fact that the ending point is nearer to the apex which is the most dynamic point of a heart, and the starting point is nearer to the aortic annulus which undergoes a relatively slower motion. In order to guarantee a safe deployment, let's assume that the registration errors take the maximum value of 2.4 mm for both the starting and ending points in either direction (i.e., total error of 4.8 mm at each side), then omitting the outer parts, the diameter of the corridor drops to 4.2 mm in systole and 17.2 mm in diastole. Since the robot always follows the centerline of the corridor, any device with diameter less than 4.2 mm can be deployed safely within such a corridor.

For all the 12 subjects, the simulated deployment of the virtual robot through dynamically registering the corridors with real-time MRI showed no collision with the inner boundaries of LV. The tested diameter of the robotic link was 4 mm. As depicted in Fig. 5 after an initial user-defined idling period, there are two deployment phases.

Table 1. The registration errors for the updated corridors on the real-time slices (n=12)

Subject/Error (mm)	End point (ε_E)			Start point (ε_S)		
	Max	Min	Average	Max	Min	Average
1	1.9	1.6	1.8	1.6	1.2	1.5
2	2.1	1.8	1.9	1.4	1.1	1.2
3	2.4	2.2	2.3	1.5	1.1	1.2
4	2.0	1.7	1.8	1.7	1.3	1.6
5	2.1	1.9	2.0	1.5	1.2	1.4
6	2.1	1.8	1.9	1.3	1.0	1.2
7	2.2	1.8	2.0	1.4	1.1	1.3
8	2.4	1.9	2.3	1.5	1.2	1.3
9	2.0	1.7	1.8	1.6	1.1	1.4
10	2.2	1.7	1.9	1.4	1.0	1.2
11	2.2	2.0	2.1	1.2	1.0	1.1
12	2.3	2.0	2.2	1.5	1.2	1.3

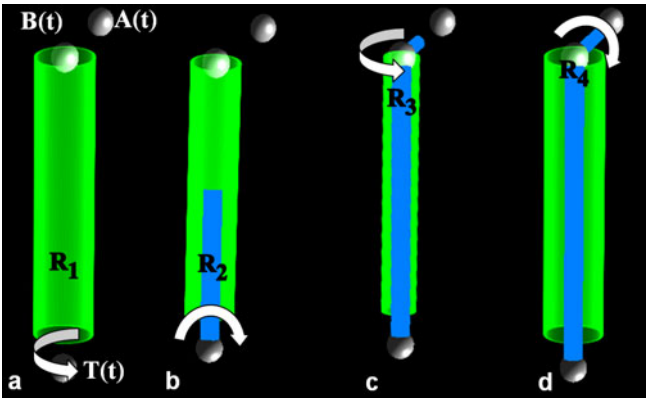


Fig. 5. Deployment is simulated (the images have the same field of view): (a) and (b) spans Phase I. Phase II starts at (c) and ends in (d).

The Phase I is the extension of the first link from the apical entrance toward the bending point, which entails the actuation of the first two rotational DOFs (i.e., R_1 and R_2) to maintain the deployed part inside the corridor. Once the distal end of the first link reaches near B(t), the Phase II starts with the extension of the second link toward the targeted aortic annulus. Concurrently, the third and fourth rotational DOFs (i.e., R_3 and R_4) are also actuated to maintain the second link along the aortic annulus midline. After reaching the target, the robot maneuvers to hold the position: the base of the robot at T(t), the robot inside the corridor, and the tip of the second link at A(t).

It is noteworthy that the presented computational core is connected to the MR scanner via a local area network (LAN) for a two-way communication: (i) real-time data transfer from the scanner to the computational core (thereby achieving a refreshing rate of 20 fps) and (ii) in reverse on-the-fly adjustment of the imaging parameters from the control module of the core, as we have demonstrated before [8].

4 Conclusions

This paper introduces a novel computational methodology for planning and performing real-time MRI-guided interventions in a beating heart. In all our studies it was able to generate a dynamic cylindrical corridor and update it with real-time single-plane MRI. Future work includes testing it on an actuated cardiac phantom [9] and automatically tracking the aortic annulus centerline on LA real-time MRI.

Acknowledgments. Supported by the National Science Foundation (NSF) award CPS-0932272. All opinions, findings, conclusions or recommendations in this work are those of the authors and do not necessarily reflect the views of our sponsors.

References

1. Jolesz, F.A.: Future perspectives for intraoperative MRI. Neurosurg. Clin. N. Am. 16, 201–213 (2005)
2. Li, M., Mazilu, D., Wood, B.J., Horvath, K.A., Kapoor, A.: A robotic assistant system for cardiac interventions under MRI guidance. In: Proc. SPIE 7625, pp. 76252X (2010)
3. McRae, M.E., Rodger, M., Bailey, B.A.: Transcatheter and transapical aortic valve replacement. Crit. Care Nurse 29, 22–37 (2009)
4. Horvath, K.A., Mazilu, D., Guttman, M., Zetts, A., Hunt, T., Li, M.: Midterm results of transapical aortic valve replacement via real-time magnetic resonance imaging guidance. J. Thorac. Cardiovasc. Surg. 139, 424–430 (2010)
5. Yeniaras, E., Deng, Z., Davies, M., Syed, M.A., Tsekos, N.V.: A Novel Virtual Reality Environment for Preoperative Planning and Simulation of Image Guided Intracardiac Surgeries with Robotic Manipulators. Stud. Health Technol. Inform. 163, 716–722 (2011)
6. Zhou, Y., Yeniaras, E., Tsiamyrtzis, P., Tsekos, N., Pavlidis, I.: Collaborative Tracking for MRI-Guided Robotic Intervention on the Beating Heart. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6363, pp. 351–358. Springer, Heidelberg (2010)
7. Elgort, D.R., Wong, E.Y., Hillenbrand, C.M., Wacker, F.K., Lewin, J.S., Duerk, J.L.: Real-time catheter tracking and adaptive imaging. J. Magn. Reson. Imaging 18, 621–626 (2003)
8. Christoforou, E., Akbudak, E., Ozcan, A., Karanikolas, M., Tsekos, N.V.: Performance of interventions with manipulator-driven real-time MR guidance: implementation and initial in vitro tests. Magn. Reson. Imaging 25, 69–77 (2007)
9. Sternberg, N.V., Hedayati, Y., Yeniaras, E., Christoforou, E., Tsekos, N.V.: Design of an actuated phantom to mimic the motion of cardiac landmarks for the study of image-guided intracardiac interventions. In: ROBIO, pp. 856–861 (2010)

Steerable Intravitreal Inserts for Drug Delivery: *In Vitro* and *Ex Vivo* Mobility Experiments*

Christos Bergeles[1], Michael P. Kummer[1], Bradley E. Kratochvil[1], Carsten Framme[2], and Bradley J. Nelson[1]

[1] Institute of Robotics and Intelligent Systems, ETH Zurich, 8092 Zurich, Switzerland
cbergeles@ethz.ch
[2] Universitätspital Bern, Inselspital, 3010 Bern, Switzerland

Abstract. The progress of wet age-related macular degeneration can now be controlled by intravitreal drug injection. This approach requires repeated injections, which could be avoided by delivering the drug to the retina. Intraocular implants are a promising solution for drug delivery near the retina. Currently, their accurate placement is challenging, and they can only be removed after a vitrectomy. In this paper, we introduce an approach for minimally invasive retinal drug delivery using magnetic intraocular inserts. We briefly discuss the electromagnetic-control system for magnetic implants and then focus on evaluating their ability to move in the vitreous humor. The mobility of magnetic intraocular implants is estimated *in vitro* with synthesized vitreous humors, and *ex vivo* with experiments on cadaver porcine eyes. Preliminary results show that with such magnetic implants a vitrectomy can be avoided.

1 Age-Related Macular Degeneration

Age-related macular degeneration (AMD) is a degenerative disorder of the posterior eye segment that is characterized by the development of lesions in the macular area. Figure 1 illustrates how the disease progresses. The result is loss in central-vision acuity.

AMD is the leading cause of blindness worldwide. In the United States, for example, it accounts for more than 50% of the cases of blindness, and, given the aging of the population, the number of affected people is expected to double [1]. Improving the efficacy of treatment methods for AMD can improve the quality of life for a significant number of patients.

2 Existing Therapies and Challenges

Current treatments for AMD aim at delivering drugs to inhibit the formation of new lesions and prevent further vision loss [2]. The most targeted drug delivery

* This work was supported by the NCCR Co-Me of the Swiss National Science Foundation.

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 33–40, 2011.

(a) (b) (c)

Fig. 1. Retinal-fundus images showing different stages of AMD: (a) AMD at an intermediate stage, (b) geographic atrophy associated with AMD, and (c) angiogenesis in AMD. (Images from the National Eye Institute, National Institutes of Health, USA).

method is intravitreal administration [3] and has been recently introduced for AMD stabilization. However, repeated injections are required for effective treatment, since the physiology of the vitreous humor reduces the amount of drug that finally reaches the macular area.

Biodegradable and non-biodegradable inserts that act as drug reservoirs are a proposed solution to this problem [4]. The reservoirs can be tailored to allow prolonged drug release, and placement of the inserts close to the retina enables targeted drug delivery. Non-biodegradable drug reservoirs are preferable due to their more controlled release mechanisms [5].

Non-biodegradable inserts must be surgically removed after their reservoir is depleted. This requires a vitrectomy, a risky invasive procedure that involves dissecting and replacing the vitreous humor with a thin solution. This requirement hinders the acceptance of the intraocular-implant treatment by patients.

3 Electromagnetically Steerable Intravitreal Inserts

Microrobotic devices are an emerging tool for minimally invasive surgery and localized drug delivery [6]. In this paper, the use of electromagnetically controlled microrobots that act as intravitreal inserts with drug reservoirs is proposed (Fig. 2(a)). Mobility allows these devices to be positioned and removed in a controllable and minimally invasive manner, potentially avoiding a vitrectomy. The microrobots are introduced in the posterior eye segment through the *pars plana* region of the sclera and are wirelessly controlled to the macula where they release their payload. Upon depletion of their reservoir they are controlled towards the entry point for removal by a magnetic tool. For AMD treatment, the microrobots can be placed in the area of the lower vessel arcade without severely affecting sight.

The microrobots have an outer diameter of less than 500 μm and fit in a 23 G needle. Their length, i.e. their reservoir, can be tailored based on the dosing required. Initially, the microrobots were microassembled soft-magnetic elliptical pieces [7], but microrobots from permanent magnetic materials are an alternative

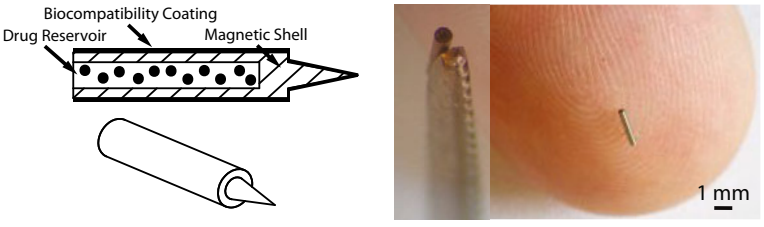


Fig. 2. (a) Concept of a magnetically driven reservoir, and (b) a 2.4 mm long and 0.38 mm wide hollow SmCo microrobot (0.135 mm inner diameter), similar to pSivida's Illuvien®

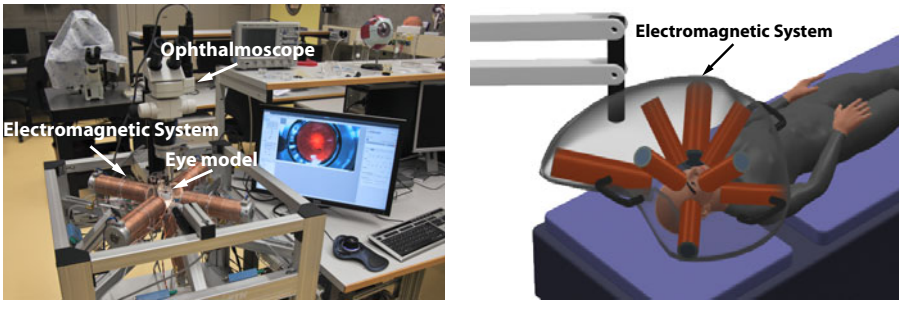
that is under examination (Fig. 2(b)). Researchers are also investigating the in-tissue mobility of screw-like devices for drug delivery [8]. Scaling these devices to the sub-mm range present fabrication challenges. Our initial experiments indicate that rotating screw-like devices get tangled up in the collagen fibers of the vitreous humor, thus, screw motion is not considered in this work.

These microrobotic devices are non-bioerrodable, and are rendered biocompatible using polypyrrole coatings [9]. Since the human eye, in general, has "immune privileges" that limit the risk of an inflammatory reaction to foreign antigens [10], severe reactions from the immune system due to the presence of coated microdevices is not expected.

4 Five Degree-of-Freedom Electromagnetic Steering

Unrestrained wireless electromagnetic control was introduced in [11], where the researchers described the OctoMag (Fig. 3), an electromagnetic control system capable of controlling magnetic devices in five degrees-of-freedom (DOF) (3-DOF position, 2-DOF pointing orientation). The workspace of the OctoMag is approximately 15 mm×15 mm and covers the posterior segment of the human eye. The system consists of eight DC-operated electromagnets arranged in a hemispherical configuration. The electromagnets are equipped with soft-magnetic cores, and, assuming that the cores are from a near-ideal material, the fields of all the magnets superimpose linearly. The introduction of soft-magnetic cores enables the generation of high field-gradients. In its current configuration, a small animal (e.g. a rabbit) can be placed between the electromagnets for *in vivo* experiments. Manual position control of the microdevices is preferred by surgeons and is enabled by the OctoMag's interface.

Currently, the OctoMag can create a maximum gradient of 1.5 $\frac{\mathrm{T}}{\mathrm{m}}$. Microdevices experience different forces when introduced to the same electromagnetic field gradient due to their different magnetic properties. Experimental estimation of the maximum applicable force indicates how they can interact with the intraocular environment.



(a) (b)

Fig. 3. (a) The OctoMag electromagnetic control system with an eye model in its workspace, and (b) rendering of a future system for AMD treatment

Table 1. Maximum force on NdFeB, and respectively, SmCo devices for various orientations

Field Orientation	$\mathbf{F}_{\text{up}}$ (μN)	$\mathbf{F}_{\text{down}}$ (μN)	$\mathbf{F}_{\text{lat,x}}$ (μN)	$\mathbf{F}_{\text{lat,y}}$ (μN)	$\mathbf{F}_{\text{lat,xy}}$ (μN)
z	287.0, 365.0	287.0, 365.0	233.9, 209.1	233.9, 209.1	287.0, 296.3
$-z$	287.0, 365.0	287.0, 365.0	233.9, 209.1	233.9, 209.1	287.0, 296.3
x	287.0, 261.4	287.0, 261.4	287.9, 387.3	287.0, 387.3	287.0, 331.1
xy	287.0, 361.2	287.0, 361.2	287.0, 369.9	287.0, 369.9	287.0, 364.1

For a magnetic field of $\|\mathbf{B}\| = 15\,\text{mT}$, the maximum generated forces on NdFeB and SmCo inserts were estimated (Table 1). The maximum force applied to soft-magnetic Ni or CoNi microrobots along every orientation is 125.7 μN [11]. The superior performance of permanent magnetic microrobots makes them more attractive for use as intravitreal drug carriers. According to [12], 75% of the forces in retinal surgery are between 0 and 7.5 mN. Being on the extreme lower range of this scale, no risk of accidental retinal penetration or tears from using permanent magnetic microrobots is expected.

5 Evaluating the Mobility of Steerable Inserts *In Vitro*

In [13], a protocol to create artificial vitreous humors with desired viscosity (loss modulus, G") and elasticity (storage modulus, G') was introduced. These artificial humors consist of water (H_2O), agar-agar (AG), and hyaluronic acid (HA). Water is 99% of the natural vitreous humor, and hyaluronic acid gives it some of its mechanical properties [14]. Agar is giving elasticity to the mixture.

Eleven fluids with different properties were synthesized (Table 2). These artificial vitreous humors cover a range from a mechanical equivalent of the porcine vitreous humor to human vitreous humor cases reported in [14]. The storage modulus for the vitreous of these human eyes is shown in Table 3. Mechanical

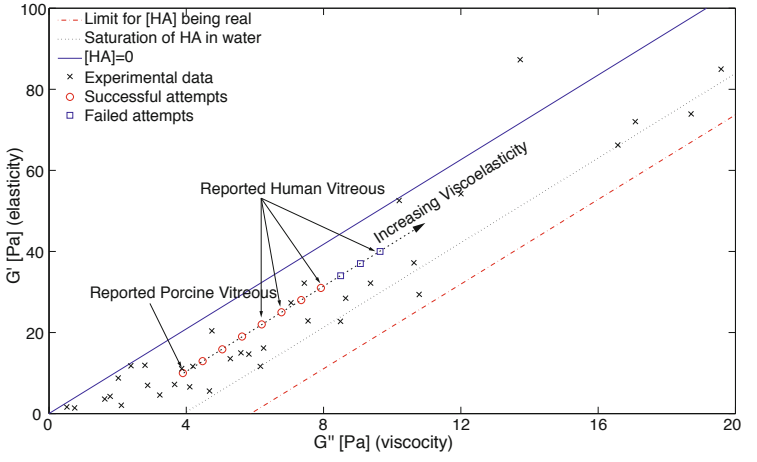


Fig. 4. Motion experiments with artificial vitreous humor. The storage and loss moduli are from [13]. The red circles indicate successful movement attempts and the blue boxes indicate that the insert was unable to translate. The crosses are the different fluids examined in [13] to extract the relationship between AG, HA, storage and loss. Image adapted from [13].

Table 2. Agar for 10 mL of water. HA amount is fixed at 2.21 mg/mL.

Fluid Case	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
AG (mg/mL)	1.098	1.27	1.42	1.56	1.68	1.80	1.91	2.014	2.11	2.21	2.30

data of the human-eye vitreous are scarce in the literature. It is known, however, that the vitreous liquifies with age [15], and this trend can also be observed in Table 3. These data enables us to synthesize fluids for experiments that provide an indication of how microdevices will behave in human vitreous humor.

According to [13], we can fix the amount of HA in solution and vary only the amount of AG to achieve desired mechanical properties. The substances are added in water heated to boiling. The solution is maintained near boiling, stirred rigorously until fully transparent, and left to cool overnight. Figure 4 shows the storage and loss moduli for different synthesized vitreous humors. Circles and squares indicate successful and failed attempts, respectively.

Experiments were conducted with 2 mm-long CoNi microrobots, and 1 mm-long NdFeB microrobots. The devices were servoed along their x-axis using the OctoMag as this orientation results in a lower fluidic resistance. We observed that even in the thinnest of the synthesized fluids, a fluid with similar viscoelastic properties to the porcine vitreous, the CoNi microrobot is unable to move. The NdFeB microrobot is able to move in a greater range of mediums, ranging from the porcine vitreous to reported human values. These human values correspond to ages 45 and 62 in Table 3, and are within the age group that the developed drug delivery method is intended for.

Table 3. Storage modulus and age for different human eyes. "L" and "R" correspond to the left and right eye respectively. Data extracted from [14].

	Patient 1		**Patient 2**		**Patient 3**		**Patient 4**	
Age	20		30		45		62	
Eye	L	R	L	R	L	R	L	R
G' [Pa]	30	27	39	5	27	25	15	3

To summarize, it is difficult to use CoNi microdevices intraocularly without a vitrectomy, but based on the experimental results, this is possible with NdFeB microrobots. This is not surprising since higher forces can be exercised on permanent-magnetic microrobots than on soft-magnetic ones.

6 Evaluating the Mobility of Steerable Inserts *Ex Vivo*

In Table 3, it can be seen that the vitreous humor of the 62 year old donor has a storage modulus similar to that of the porcine vitreous humor. Additionally, a study performed in [16] reports that the central vitreous humor of the human and the pig are rheologically similar. The same study states that the anterior and posterior regions of the porcine vitreous resemble a "thick" gel, whereas the human vitreous has a "thinner" and in some cases "watery" consistency. We can conclude that, with respect to the properties of the vitreous humor, porcine eyes can be used as a "hard" model of the human eye.

Studies on the human vitreous humor are conducted with eyes from elderly donors, whereas porcine eyes are taken from relatively young animals. This fact is also acknowledged in [16]. From personal communication with vitreoretinal surgeons, it was understood that successful mobility of microrobots in porcine eyes received from slaughterhouses will be a good indicator for similar or better mobility performance in the eyes of elderly human patients.

Thus, the goal is to evaluate the mobility of the microdevices in porcine vitreous humor. In [14], it is stressed that the vitreous humor rapidly loses hyaluronic acid upon removal from the eye globe, it collapses, and its mechanical properties change significantly. Thus, mobility experiments need to be performed in the eye globe. As in [12], the cadaver eye is dissected above the *pars plana* region, and the anterior chamber and iris are exposed after removing the cornea. Subsequently, the intraocular lens and iris are removed, and the vitreous humor is accessible. Removing the optical elements facilitates visualization, and no ophthalmoscope is required for observation.

The microrobot is inserted in the vitreous humor using a syringe equipped with a glass pipette. Afterwards, the eye is positioned under the OctoMag for experimentation. The eye is illuminated transsclerally, and the images are captured by a Leica M80 microscope equipped with a Grasshopper-14S5M/C camera.

Results for a NdFeB microrobot moving in the vitreous humor of two cadaver porcine eyes are shown in Fig. 5. The mobility of the microrobots supports the

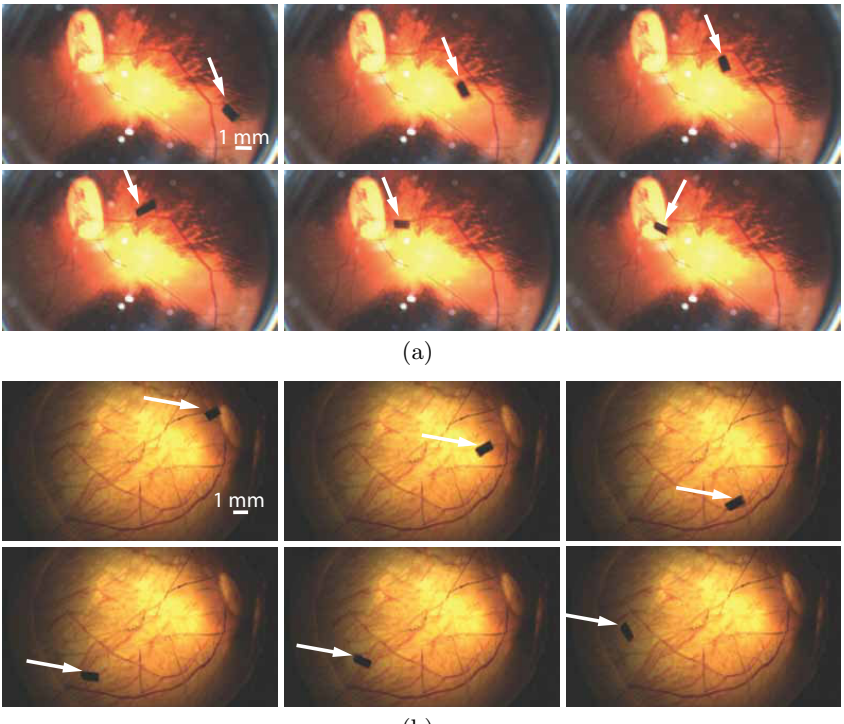


Fig. 5. Sequences of frames showing a 1 mm long NdFeB microrobot moving in the vitreous humor of two cadaver porcine eyes

in vitro experimental results and suggests that magnetic microrobotic inserts can be used in human eyes without necessitating a vitrectomy. This, together with the increased mobility the electromagnetic control offers, will enable the minimally invasive delivery of existing drugs closer to the retina.

7 Conclusions and Future Work

In this paper, the concept of performing minimally invasive localized drug delivery in the posterior eye segment using steerable microrobotic drug carriers was introduced. The aim of the current technology is to assist in the treatment of age-related macular degeneration, however, it can also be used for other posterior-eye diseases. Magnetic microrobots can be considered a safe tool for localized drug delivery. *In vitro* experimentation demonstrated that permanent-magnetic microrobots are able to move in a range of fluids covering porcine vitreous and human vitreous. Successful *ex vivo* experiments lead to the conclusion that magnetic microrobots have the potential to be wirelessly controlled in human eyes without a vitrectomy, making drug delivery in the posterior-eye segment more targeted and less invasive. Future work will include a statistical analysis of repeated experiments and design optimization of the microdevices.

References

1. Congdon, N., O'Colmain, B., Klaver, C.C., Klein, R., Munoz, B., Friedman, D.S., Kempen, J., Taylor, H.R., Mitchell, P.: Causes and prevalence of visual impairment among adults in the United States. Arch. Ophthalmology 122(4), 477–485 (2004)
2. Nowak, J.Z.: Age-related macular degeneration: pathogenesis and therapy. Pharmacological Reports 58(3), 353–363 (2006)
3. Lee, S.S., Robinson, M.R.: Novel drug delivery systems for retinal diseases. Ophthalmic Research 41(3), 124–135 (2009)
4. Danckwerts, M.P., Fassihi, A.: Implantable controlled release drug delivery systems. Drug Development and Industrial Pharmacy 17(11), 1465–1502 (1991)
5. Yasukawa, T., Ogura, Y., Tabata, Y., Kimura, H., Wiedemann, P., Honda, Y.: Drug delivery systems for vitreoretinal diseases. Progress in Retinal and Eye Research 23(3), 253–281 (2004)
6. Nelson, B.J., Kaliakatsos, I.K., Abbott, J.J.: Microrobots for minimally invasive medicine. Annual Review of Biomedical Engineering 12, 55–85 (2010)
7. Yeşin, K.B., Vollmers, K., Nelson, B.J.: Modeling and control of untethered biomicrorobots in a fluidic environment using electromagnetic fields. Int. J. Robotics Research 25(5-6), 527–536 (2006)
8. Ishiyama, K., Sendoh, M., Yamazaki, A., Arai, K.I.: Swimming micro-machine driven by magnetic torque. Sensor and Actuators A: Physical 91(1–2), 141–144 (2001)
9. Sivaraman, K.M., Bayrakceken, K., Ergeneman, O., Pane, S., Lühmann, T., Hall, H., Nelson, B.J.: Tailoring the drug loading capacity of polypyrrole films for use in intraocular biomicrorobots. In: IEEE Int. Conf. Engineering in Medicine and Biology, pp. 4359–4362 (2010)
10. Choonara, Y.E., Pillay, V., Danckwerts, M.P., Carmichael, T.R., Du Toit, L.C.: A review of implantable intravitreal drug delivery technologies for the treatment of posterior segment eye diseases. J. Pharmaceutical Sciences 99(5), 2219–2239 (2010)
11. Kummer, M.P., Abbott, J.J., Kratochvil, B.E., Borer, R., Sengul, A., Nelson, B.J.: OctoMag: An electromagnetic system for 5-DOF wireless micromanipulation. IEEE Trans. Robotics 26(6), 1006–1017 (2010)
12. Gupta, P.K., Jensen, P.S., de Juan Jr., E.: Surgical forces and tactile perception during retinal microsurgery. In: Taylor, C., Colchester, A. (eds.) MICCAI 1999. LNCS, vol. 1679, pp. 1218–1225. Springer, Heidelberg (1999)
13. Kummer, M.P., Abbott, J.J., Dinser, S., Nelson, B.J.: Artificial vitreous humor for in vitro experiments. In: IEEE Int. Conf. Engineering in Medicine and Biology, pp. 6406–6409 (2007)
14. Nickerson, C.S.: Engineering the mechanical properties of ocular tissues. PhD thesis, California Institute of Technology (2005)
15. Los, L.I., Van der Worp, R.J., Van Luyn, M.J.A., Hooymans, J.M.M.: Age-related liquefaction of the human vitreous body: LM and TEM evaluation of the role of proteoglycans and collagen. Investigative Ophthalmology and Visual Science 44(7), 2828–2833 (2003)
16. Lee, B., Litt, M., Buchsbaum, G.: Rheology of the vitreous body: Part 2. viscoelasticity of bovine and porcine vitreous. Biorheology 31(4), 327–338 (1994)

Distance Measurement in Middle Ear Surgery Using a Telemanipulator

Thomas Maier[1], Gero Strauss[2], Franz Bauer[1], Andreas Grasser[1], Nobuhiko Hata[3], and Tim C. Lueth[1]

[1] Institute of Microtechnology and Medical Device Technology (MiMed), Technische Universität München, Germany
thomas.maier@tum.de
[2] Department of ENT, University Hospital Leipzig and BMBF Innovation Center Computer Assisted Surgery (ICCAS), University Leipzig, Germany
gero.strauss@medizin.uni-leipzig.de
[3] Department of Radiology, Harvard Medical School, Brigham and Women's Hospital, USA
hata@bwh.harvard.edu

Abstract. In this article, a new tool for the intraoperative measurement of distances within the middle ear by means of a micromanipulator is presented. The purpose of this work was to offer the surgeon a highly accurate tool for measuring the distances between two points in the 3D operational field. The tool can be useful in various operations; this article focuses, however, on measuring the distance between the stapes footplate and the long process of the incus of the middle ear. This distance is important for estimating the proper prosthesis length in stapedotomy for treating otosclerosis. We evaluated the system using a simplified mechanical model. Our results show that the system can measure distances with a maximum error of 0.04 mm.

Keywords: Middle Ear Surgery, Telemanipulator, Stapedotomy, Measurement.

1 Introduction

The intraoperative measuring of anatomical structures is necessary in many surgical interventions, in order to evaluate preoperative situations and postoperative results. With a large surgical access, this can be done easily. In the small operational field of Middle Ear Surgery, however, this is much more difficult [1]. An example is the measurement of the distance between the long process of the incus and stapes footplate in a stapedotomy. Stapedotomy is a surgical technique for treating conductive hearing loss owing because of ossification of the oval window. Through stapedotomy, the stapes is replaced by a small prosthesis (Fig. 1 a, b). The prosthesis should protrude about 0.2 - 0.3 mm into the inner ear [2]. Depending on the patient's anatomy, prostheses of different length (e.g., 3.25 mm – 4.7 mm) are used.

The prosthesis length has a great impact on the hearing results. A prosthesis that is too long can lead to dizziness, sudden pressure gains, and conductive hearing loss. A prosthesis that is too short, however, can lead to a displacement of the prosthesis. Furthermore, the perforation in the stapes footplate can be overgrown, which

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 41–48, 2011.

subsequently leads to an interception of sound conduction. Results can be conductive hearing loss and perilymph fistula [4, 2]. A revision surgery, in which the incorrect prosthesis is replaced by a suitable one, can be necessary. According to literature, 7-28% of the stapes revision surgeries can be directly traced back to incorrect prosthesis length [4, 3]. Gros et al. even indicates that incorrect prosthesis length is a major reason for prosthesis dislocations, which are the cause of 18-82% of the revisions [4].

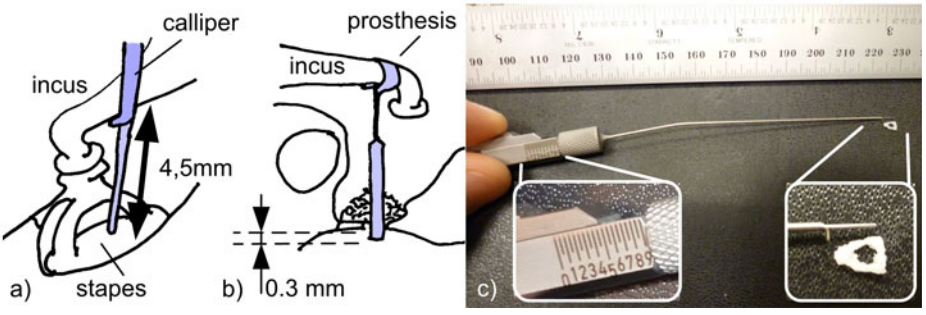


Fig. 1. a) Distance between incus and stapes footplate. b) Stapes replaced by prosthesis. Calliper with stapes model (Heinz Kurz GmbH Medizintechnik, Dusslingen, Germany).

In clinical practice, intraoperative measurements of anatomical structures during middle ear surgeries are performed with special callipers (Fig. 1a, c), measuring rods, sizers or through estimation based on anatomical landmarks [1]. Most often, a prosthesis length of 4.5 mm is required. Hence, some surgeons use only this single prosthesis type or try to estimate the proper length. Kwok et al. assume that it is impossible to do an accurate estimation of the distance between the long process of the incus and the stapes footplate without measuring devices, owing to the angle of view. Furthermore, Strauss et al [1] describe the magnification and the low depth of field caused by the surgical microscope as difficult for estimations. In [5], optical coherence tomography (OCT) is used to determine the proper prosthesis length. Moreover, the OCT measurements are compared with those performed with a mechanical calliper. It is found that measuring with a calliper leads to an error of up to 0.4 mm. In research, the positioning of prostheses was analyzed postoperatively with high-resolution CT in [6]. It was found that HR-CT data is not suitable for doing accurate measurements in this scale. In [1], a new optical multipoint laser-measurement system (KARL STORZ GmbH & Co. KG, Tuttlingen, Germany) for a flexible endoscope is introduced for ENT-surgery. The accuracy is reported to be 0.2 to 0.5 mm.

The distance measurement between stapes footplate and long process of the incus during stapedotomy involves strict requirements in terms of the accuracy and usability of the measuring system. Hence, this measurement was chosen representatively for many other measurement tasks in ENT-Surgery. The system should provide measurements with accuracy better than 0.1 mm within the operational field. We propose a telemanipulated measurement system with 3 degrees of freedom (DOF) and highly accurate position sensors. By touching two anatomical landmarks using arbitrary microinstruments like perforators or forceps, their relative distance should be

calculated and visualized. Measurements should be able to be taken in stand-alone operation of the manipulator or in combination with external devices, such as navigation systems or systems for image processing.

2 Description of the Measuring System

The measuring system used consists of a teleoperated micromanipulator with joystick control, three highly accurate position sensors, an independent microcontroller-based evaluation electronics, a visualization software, and a bridge software for protocol transformation (Fig. 2).

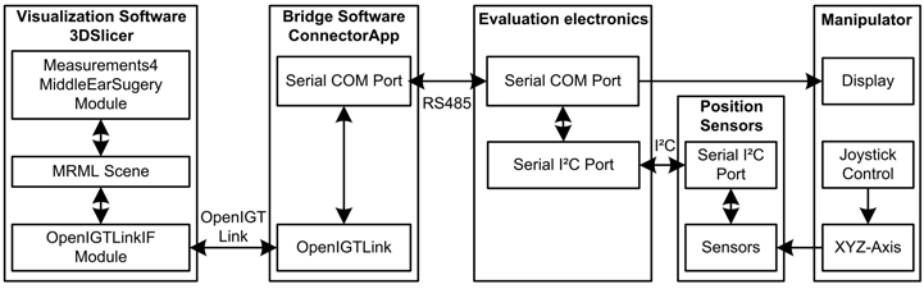


Fig. 2. Block diagram of the overall measuring system

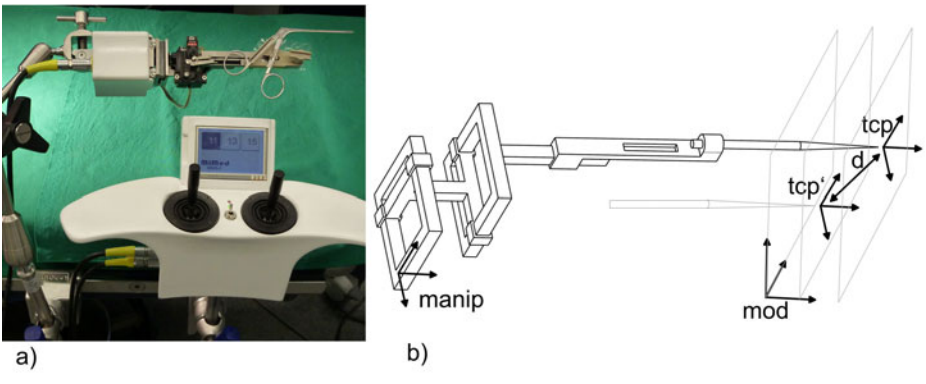


Fig. 3. a) Micromanipulator System with joystick control. b) Cartesian kinematics of the Manipulator with 3DOF. The distance between two tcp-positions can be calculated within the CT model.

Micromanipulator: A Micromanipulator System developed for middle ear surgery was introduced in [7] (Fig. 3). This system is based on previous experience with large ceiling-mounted interactive robots [8] and small autoclavable robots [9-10]. We used this manipulator as the basis for our measuring system. It is small-sized, lightweight, and capable of moving standard instruments of ENT surgery with high accuracy in a

small workspace. It can be already be integrated into standard operating procedures and allows rapid change between manual and teleoperated instrument guidance. The micromanipulator has three translational DOF in Cartesian configuration. An additional DOF is used to open and close attached forceps. The MMS-II is mounted on the OR table's side rail and aligned with the operating field, using an articulated arm. The joystick controller is mounted on the OR table's side rail directly in front of the surgeon.

Measuring System and Sensors: Each axis of the manipulator was equipped with an additional position sensor. Because of their high precision, small overall size, and simple interface, hall-sensors of the type TRK-1T02 Tracker (New Scale Technologies Inc., Victor, NY, USA) were chosen. They are able to detect an absolute position within a terminal pair (2 mm) with 12-bit resolution. By detecting the crossover from one terminal pair to the next, we are able to cover an arbitrary measuring section. Positions can be read out via I^2C serial bus. The strength and gradient of the magnetic input are analyzed in order to check the integrity of the data obtained. According to the manufacturer, the maximum absolute error is ±10 µm.

Evaluation Electronics: A microcontroller-based evaluation electronics was built to readout the sensors. An Atmega 168 microcontroller (Atmel Corporation, San Jose, CA, USA) with 20 MHz and 8 bit is used as a central processing unit. The controller features an I²C interface for communicating with the position sensors. Furthermore, the evaluation electronics has an interface for sending the current position continuously (20 Hz) to external components and devices via RS232 or RS485. This is used to visualize the manipulator at an external navigation system. Additionally, the current position can be sent and saved as a measuring value. Therefore, the sensor readings are averaged within a time interval of 0.5 seconds to increase accuracy. The averaged position is sent and stored as a measuring value only when the sensor readings have been stable in that time interval. The communication protocol used is proprietary to our lab and enables communication with other robotic and image-processing devices. This ensures modularity. The Euclidian distance between two measuring values is also calculated directly by the microcontroller in order to be displayed at the joystick control, if stand-alone operation without navigation is needed.

Visualization Software: To visualize the measuring values and results, the open-source software 3DSlicer (www.slicer.org) was used. 3DSlicer is a cross-platform application for visualizing and analyzing medical-image data. With the help of various modules, image and sensor data can be recorded, processed, and displayed. Various data (images, positions, transforms, etc.) can be exchanged with external devices using the OpenIGTLink protocol [11]. We used the data type TRANSFORM to exchange data with 3DSlicer. The evaluation electronics provides the position data and measurement values using our laboratory's own proprietary protocol, which is not compatible to OpenIGTLink. Therefore, a bridge software (ConnectorApp) was implemented. This software transforms message data received from the proprietary protocol to the OpenIGTLink protocol. The ConnectorApp has the additional task of coordinating the different working steps (initialization, registration, and measurement). It further transforms positions from manipulator coordinates to 3DSlicer image coordinates. The calculation of the distance between two measuring values and the

visualization of those points and their distance are carried out by the newly developed "Measurements4MiddleEarSurgery" (MMES) module for 3DSlicer.

Initialization and Measuring Process: The system usage introduced is divided into three steps: initialization, registration, and measurement. The initialization step sets up communication between 3DSlicer, ConnectorApp, and evaluation electronics. For that purpose, the 3DSlicer module "OpenIGTLinkIF" is used. After that, basic information is interchanged. This is necessary to define the way data is exchanged in the next steps. During the registration step, the transformation matrix between 3DSlicer's manipulator coordinates (manip) and image coordinates (mod) must be calculated. Therefore, three landmarks are set in the image dataset loaded within 3DSlicer. This is done by means of the "Fiducials" module. Possible landmarks are bony structures within the middle ear, e.g. the bony canal of the facial nerve. After that, the three points are sent to the ConnectorApp using OpenIGTLink. The same landmarks are then set with the manipulator's tooltip and sent to the ConnectorApp, which is then able to calculate the necessary transformation matrix. This step can be skipped if the visualization of points within the 3D image dataset is not required. Finally, the third step--the measurement--can be started. The evaluation electronics starts transmitting the current position every 50 ms. The ConnectorApp performs the coordinate transformation and routes the data via OpenIGTLink to 3DSlicer. 3DSlicer receives the data within the OpenIGTLinkIF module. The position of the visualized tool tip can now be updated. To record a measurement value, a command can be sent to the evaluation electronics by a simple switch or using the serial port. As described above, the sensor readings are averaged; their correctness is tested; and the resulting measurement value is sent to 3DSlicer. As soon as a second measurement value is received by 3DSlicer, the distance between those two is calculated and visualized.

3 Experiments and Results

Two Experiments were conducted to validate the measuring system. In the first experiment, the accuracy itself of the sensor system was determined. In the second experiment, the telemanipulated instrument guidance was considered in the accuracy measurement.

Experiment 1: Accuracy of the Sensor System
In this experiment, the accuracy of the sensor system was valuated. Therefore, the thickness of two precise gauge blocks was determined. The manipulator motors were deactivated, so that the manipulator's precision would not affect the measurement results. Five measuring points were marked on a measuring table within an area of 15 x 15 mm². The Manipulator was positioned in such a way that the axis of the inserted instrument (STORZ Perforator, 0.3 mm) was oriented perpendicular to the measuring table. The instrument tip was moved manually toward the first measuring point. A first measurement value was recorded at that position. After that, the instrument tip was moved upward. Two precision gauge blocks, with a thickness of 2 mm (± 0.01 mm) and 5 mm (± 0.01 mm) respectively, were put onto the measuring table. Then, the instrument tip was set down again, and this position was stored as the second

measuring value. The calculated distances, shown by 3DSlicer, were stored. This procedure was repeated twenty times for each of the five measuring points. The mean and the standard deviation (SD) are shown in Table 1.

Table 1. At five measuring points, the thickness of two gauge blocks was determined

	Measuring Point	**1**	**2**	**3**	**4**	**5**
gauge block: 2.00 mm ± 0.01 mm	**mean/mm (n=20)**	2.00	2.00	2.00	2.00	2.00
	SD/mm	0.01	0.01	0.01	0.01	0.01
	maximum error			0.03		
gauge block: 5.00 mm ± 0.01 mm	**mean/mm (n=20)**	5.00	4.99	4.99	5.01	5.00
	SD/mm	0.01	0.01	0.01	0.01	0.01
	maximum error			0.02		

Experiment 2: Accuracy of the telemanipulated measuring system

In the second experiment, the entire measuring system, including the telemanipulated positioning of the tool tip, was valuated. A phantom was developed to imitate the task of measuring the distance between incus and stapes footplate (Fig. 4). It consists of a rod (Ø 1 mm) fastened to a cylinder by means of two elastic bands. This imitates the movable incus. The cylinder has an inner diameter of 7 mm, representing the ear canal. In the initial position, the distance between the rod and the bottom of the cylinder is 4.30 mm (±10 µm). A 1.5-mm microhook was clipped onto the manipulator. The instrument and the phantom were connected with simple electronics signaling the physical contact between both, by means of an LED. This was done in order to eliminate operator and visual-condition influence.

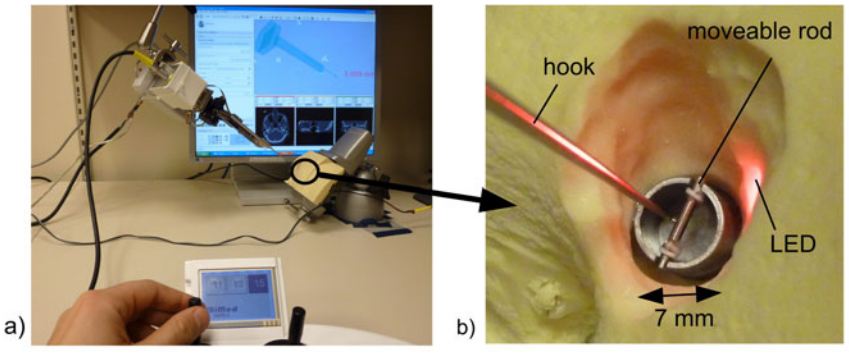


Fig. 4. a) Experimental setup. b) Detailed view of the middle-ear phantom.

The Manipulator was oriented in a way that allowed the instrument axis to be perpendicular to the cylinder ground. The instrument was moved toward the bottom (stapes footplate), using the joystick control, until the LED lighted up. The measuring value was sent to 3DSlicer. Then, the hook was moved toward the bottom side of the

rod (proximal side of the incus) until the LED lighted up. This measuring value was sent to 3DSlicer as well. The measuring result displayed by 3DSlicer was stored. Because of the moveable rod, the positioning errors of the manipulator system are included in the measuring result, in contrast to the first experiment. The results are shown in Table 2. The 0.30-mm thickness of the hook itself has to be added to the measuring value.

Table 2. Results for telemanipulated measuring

reference value:	**mean/mm (n = 50)**	4.01 mm (+ 0.30 mm hook)
4.30 mm ± 0.01mm	**SD/mm**	0.02 mm
	maximum error	0.04 mm

Results: The first experiment shows that the position sensors can determine the tool-tip position with a precision (two standard deviations) of ±0.02 mm. The small deviation of the mean value shows that there is no systematic error. The second experiment shows that the manipulator is able to touch a surface so precisely that the high accuracy of the sensors can be used in practical applications.

4 Conclusion

We have developed a measuring system based on a miniature telemanipulator for precise determination of distances within a small workspace. The aim of our work was to increase measuring accuracy in middle ear surgery. This article was focused on one specific task representing many others: the estimation of the distance between the stapes footplate and the long process of the incus during a stapedotomy. An available micromanipulator with 3DOF was equipped with an independent, highly accurate position-sensor system. The open-source software 3DSlicer (www.slicer.org) was used for visualizing the measuring results. The measuring system can also be operated without visualization and navigation in a stand-alone mode. Precise measuring is crucial in many surgical procedures. Measurements accurate to a tenth, however – as required in middle-ear surgery – cannot be achieved with common methods. Our solution can provide measurements with a maximum error of 0.04 mm, which was valuated with a technical phantom. Preliminary experiments show, that the contact of a tooltip with the tissue's surface can be determined very precisely (<0,05mm), when using a microscope. Nevertheless, this in vivo precision has to be determined in further experiments with a larger group of surgeons, because it could depend on individual skills.

Acknowledgement. The authors wish to thank Sam Song and Junichi Tokuda for their support. We also wish to thank KARL STORZ and Heinz KURZ for supporting us with medical instruments. This work has been supported by the German Research Foundation (DFG, PAK 404) and several NIH funds (5U41RR019703, 5P01CA067165, 1R01CA138586, 1R01CA124377, 2R44DE019322).

References

1. Strauss, G., Hipp, S., Fischer, M., Hofer, M., Munz, F., Dietz, A., Strauss, M., Lüth, T.: Intraoperative volumetry in ENT surgery. Objectivation of Surgical Success by the Volume Control System. HNO 58, 839–845 (2010)
2. Theissing, J., Rettinger, G., Werner, J.A.: ENT-Head and Neck Surgery: Essential Procedures. Thieme Medical Pub. (2010)
3. Edelstein, D.R.: Revision Stapedectomy and Stapedotomy. Thieme (2008)
4. Gros, A., Vatovec, J., Žargi, M., Jenko, K.: Success rate in revision stapes surgery for otosclerosis. Otology & Neurotology 26, 1143–1148 (2005)
5. Heermann, R., Hauger, C., Issing, P.R., Lenarz, T.: Application of Optical Coherence Tomography (OCT) in Middle Ear Surgery. Laryngo-Rhino-Otologie 81(6), 400–405 (2002)
6. Bozzato, A., Struffert, T., Hertel, V., Iro, H., Hornung, J.: Analysis of the accuracy of high-resolution computed tomography techniques for the measurement of stapes prostheses. European Radiology 20, 566–571 (2010)
7. Maier, T., Strauss, G., Hofer, M., Kraus, T., Runge, A., Stenzel, R., Berger, T., Dietz, A., Lueth, T.C.: A New Micromanipulator System for Middle Ear Surgery. In: IEEE International Conference on Robotics and Automation, pp. 1568–1573 (2010)
8. Hein, A., Lueth, T.C.: Image-Based Control of Interactive Robotics Systems. In: Taylor, C., Colchester, A. (eds.) MICCAI 1999. LNCS, vol. 1679, pp. 1125–1132. Springer, Heidelberg (1999)
9. Schauer, D., Hein, A., Lüth, T.C.: Dynamic force control for a miniaturised medical robot system. In: Proceedings 2003 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM 2003), pp. 1090–1095. IEEE, Port Island (2003)
10. Koulechov, K., Rapoport, T., Lueth, T.C.: Miniaturized, Autoclavable Surgical Robot (Miniaturisierter, autoklavierbarer Roboter für die Chirurgie). at - Automatisierungstechnik 54, 213–221 (2006)
11. Tokuda, J., Fischer, G.S., Papademetris, X., Yaniv, Z., Ibanez, L., Cheng, P., Liu, H., Blevins, J., Arata, J., Golby, A.J., Kapur, T., Pieper, S., Burdette, E.C., Fichtinger, G., Tempany, C.M., Hata, N.: OpenIGTLink: an open network protocol for image-guided therapy environment. The International Journal of Medical Robotics and Computer Assisted Surgery 5, 423–434 (2009)

An Instantiability Index for Intra-operative Tracking of 3D Anatomy and Interventional Devices

Su-Lin Lee[1], Celia Riga[2,3], Lisa Crowie[2], Mohamad Hamady[2], Nick Cheshire[2,3], and Guang-Zhong Yang[1]

[1] The Hamlyn Centre for Robotic Surgery
[2] Regional Vascular Unit, St Mary's Hospital
[3] Academic Division of Surgery,
Imperial College London, UK
{su-lin.lee,g.z.yang}@imperial.ac.uk

Abstract. The tracking of tissue deformation, imaging probes and surgical instruments is an integral part of intra-operative surgical guidance. While the recent introduction of shape instantiation provides a systematic framework for tracking 3D anatomy in real-time, deviations to the desired imaging location can affect the accuracy of the predicted shape. To examine the sensitivity of the prescribed image planes to such errors, we introduce in this paper a new Instantiability Index for providing the intrinsic level of robustness while using such scan planes for the tracking of anatomy and interventional devices. Optimisation of the Index is applied to 3D anatomical reconstruction and the localisation of an intraoperative imaging device. Results are shown on detailed phantom experiments for both real-time 3D shape instantiation and imaging catheter tracking.

Keywords: instantiability index, shape instantiation, tracking, localization.

1 Introduction

For intraoperative surgical guidance, effective tracking of tissue deformation and interventional devices is an integral part of the workflow. Despite increasing advances of intra-operative imaging techniques, particularly intra-operative MR and CT, the acquisition of real-time 3D sequences is often limited by practical constraints such as temporal resolution and potential radiation burden. For anatomical structures undergoing large tissue deformation, instantiating the entire dynamic 3D shape from limited imaging planes or sparsely sampled data becomes a pertinent research topic. The recent introduction of shape instantiation reduces the amount of real-time imaging data required to reconstruct an entire 3D shape by using models built preoperatively. For example, the work of Lee *et al.* [1] proposed the use of optimal scan planning for the instantiation of the organ of interest using partial least squares regression during surgery. The work balances the information content of an individual or a set of image planes against practical constraints such as available acquisition windows for capturing rapid shape deformation or transit of a contrast bolus. For intra-vascular intervention, this approach is advantageous due to a reduction in both

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 49–56, 2011.

the use of contrast agents to minimise nephrotoxicity and exposure to radiation. In the existing work reported thus far, the issue of errors in localisation of the scan plane, however, has not been explicitly addressed. Despite the availability of the optimal scan planes, in practice, it is not always possible to accurately localise to these plans due to co-registration errors or patient motion. This difficulty is particularly evident with handheld imaging probes such as ultrasound. With the original framework of shape instantiation, it is assumed that the captured scan plane will conform to the optimal one defined. Deviation to the prescribed optimal scan plane(s) or errors induced due to anatomical access difficulties can lead to significant errors in shape instantiation. It is therefore desirable to consider not only the information content of a given image plane in terms of its shape instantiability, but also the underlying resilience to perturbations or localisation errors. This situation is illustrated schematically in Fig. 1 where the robustness of two different scan planes is shown via their shape instantiation results.

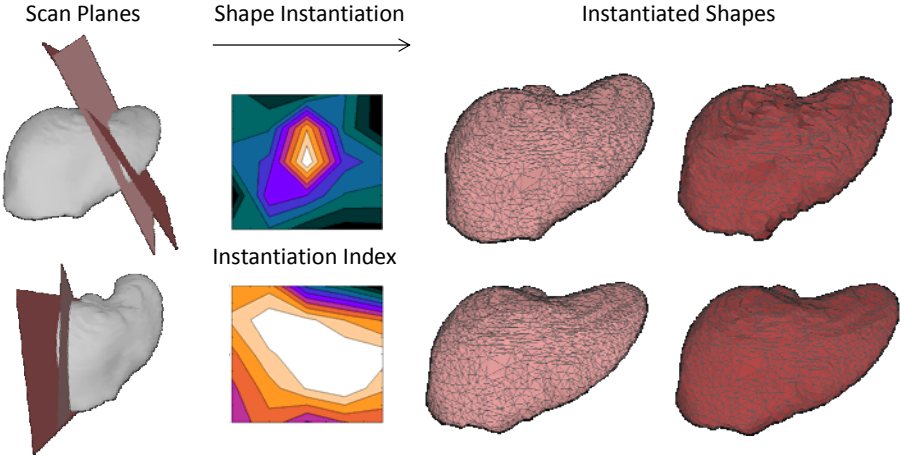


Fig. 1. The use of the Instantiability Index with respect to the existing shape instantiation framework, here with a liver example. An examination of the instantiation index for a scan plane indicates how a perturbation in its position affects the instantiation: (top) a scan plane that is less and (bottom) one that is more robust.

The purpose of this paper is to introduce a new shape Instantiability Index for accurate and more robust intra-operative 3D shape instantiation. We will demonstrate its practical use for both 3D anatomical reconstruction and reliable tracking of interventional devices. For the former, we will illustrate the underlying reasons of introducing the shape instantiability index and how it reflects the potential errors in shape instantiation caused by deviations/perturbations from the defined scan planes whereas for the latter, we demonstrate its novel use for the tracking of IVUS catheters during intravascular interventions. The tracking of IVUS catheters is known to be difficult without the use of additional positional sensors or repeated contrast-enhanced x-ray guidance and methods based on unscented particle filters [2] and image based

co-registration [3] have been proposed. The method introduced in this paper represents an alternative approach that allows simultaneous geometrical recovery and device tracking.

For detailed quantitative evaluation, the use of the Index to optimise the position of the scan plane for shape instantiation was demonstrated with a deformable silicone liver phantom. The scan plane was optimised, used to reinitialise the framework, and used to instantiate the liver model. The Index was also used to solve the inverse problem of finding the position of an IVUS catheter *in situ*, as demonstrated with a silicone arterial model.

2 Methods

2.1 Instantiability Index for Shape Recovery

The basic concept of shape instantiation can be described as follows. Prior to an operation, a statistical shape model can be built from preoperative images and an optimal scan plane determined, along which the most information of the entire shape is derived. The model is trained (*e.g.*, through the use of partial least squares regression) with this data and further image planes obtained intraoperatively are input to the regression which predicts the entire organ shape in real-time.

The Instantiability Index is based on the comparison of the real-time image data to the contours of the preoperative model along the scan plane. An examination of the Index for the scan plane during perturbation gives an indication of the sensitivity of the plane, *i.e.*, how prone to errors a small perturbation to the plane parameters can cause. A stable Index indicates a robust scan plane, with a small error in the plane orientation or position not affecting the shape instantiation. A sharp peak in the Index typically indicates a unique solution but the corresponding scan plane will be sensitive to error. Albeit undesirable for shape instantiation, this location can be used to localise the position of an *in vivo* imaging probe, as it suggests the local geometry is unique.

With the proposed framework, contours P are found by the intersection of the scan plane with the preoperative shape model, where the plane p is defined with an origin o and a normal n. Geometrically, the preoperative surface mesh consists of triangles defining the geometry of the organ of interest. The contours are then projected to 2D for comparison to the real-time imaging data. With two 3D vectors x and y, from the image origin o, along the respective axes of the 2D image, the 2D coordinates of any 3D point a are:

$$(u, v) = \left\{\left(a - o\right) \cdot x, \left(a - o\right) \cdot y\right\}. \tag{1}$$

Preprocessing of the ultrasound images was performed prior to optimisation. The objective function to be minimised, defined as the Instantiability Index (*II*), is then

$$II = f\left(R, T\right) = -\sum_{i=0}^{N}\left(I_i \times \left(RP_i + T\right)\right) \tag{2}$$

where I_i is the intensity of the image at position i, R and T are the rotation and translation of the scan plane (the parameters to be optimised), and P is the intersection between the scan plane and the preoperative model (with N points in the contour). A comparison of optimisation methods was performed with Simulated Annealing [4] performing most reliably. With the optimised plane, the partial least squares regression was reinitialised semi-automatically; the preoperative statistical shape model was recut and the points on the contour resampled. This updated partial least squares regression was used to predict the shape.

2.2 Data Collection

A number of calibration techniques and corresponding phantoms have been developed for freehand 3D ultrasound systems [5]. For this work, a three-point crossed-wires phantom was built for the calibration of intra-operative ultrasound images (coordinate system U) to the coordinate space of a tracking device (coordinate system N). Likewise, the transforms were also found which mapped the coordinate space of an electromagnetic tracker to that of the CT imaging space (coordinate system C), calculated using PRAXIS [6]. This then results in a vector defining translation and a quaternion for rotation between the ultrasound image points to the original CT imaging space. Using the notation from Prager *et al.*[7], this results in the following calculation: ${}^{C}x = {}^{C}T_{N}\,{}^{N}T_{U}\,{}^{U}x$, where ${}^{C}T_{N}$ is the transformation from the coordinate system C to N and ${}^{N}T_{U}$ is that from U to N.

For detailed validation of the proposed method, ultrasound images of a silicone organ phantom [8] were collected; this near life-sized phantom consists of a number of silicone organs, including modified lungs with inserts to simulate respiration and hence deform the silicone liver. Translation of the liver was not achieved. The data were collected on an ALOKA prosound α10 system (Aloka Co. Ltd, Tokyo, Japan). A 2D convex probe, with a custom built mount for a passive infrared marker to be tracked by an NDI Polaris tracker (Northern Digital, Inc, Waterloo, ON, Canada), was used. Calibration between the two coordinate systems is described in the next section. The ultrasound images were captured using a PC video capture card connected to the S-video output feed of the scanner. Scans were obtained at 7 different liver positions.

To demonstrate the use of the proposed instantiability index for tracking *in vivo* imaging probes, IVUS images were acquired on a Volcano IVUS system (Volcano, San Diego, CA, USA) at an anonymous hospital. An Elastrat (Elastrat Sarl, Geneva, Switzerland) pulsatile silicone phantom of the aortic arch was imaged with IVUS and with a corresponding CT roll scan with a Philips Allura Xper CT scanner. A single fluoroscopic image was also captured to provide an initial position for the IVUS catheter tip; as both scans were obtained on the same scanner, no explicit registration was required. Ten catheter positions were collected. Analyze (AnalyzeDirect, Inc, Overland Park, KS, USA) was used to perform the segmentation and model building. For the IVUS data, the position of the tip of the catheter was initialised by obtaining its rough position in two dimensions from the corresponding fluoroscopic image. The optimisation of the instantiability index was then used to determine the final position of the catheter within the mesh.

3 Results

3.1 Validation for Shape Instantiation

For robust anatomical shape instantiation with data collected from the Aloka scanner, the scan plane was first optimised with respect to the preoperative model. With this optimised scan plane, partial least squares regression for the shape instantiation was reinitialised. The errors were calculated as the mean of the distances from the points on the instantiated shape to the closest points on the ground truth model, the 3D mesh of the liver as manually segmented from CT images.

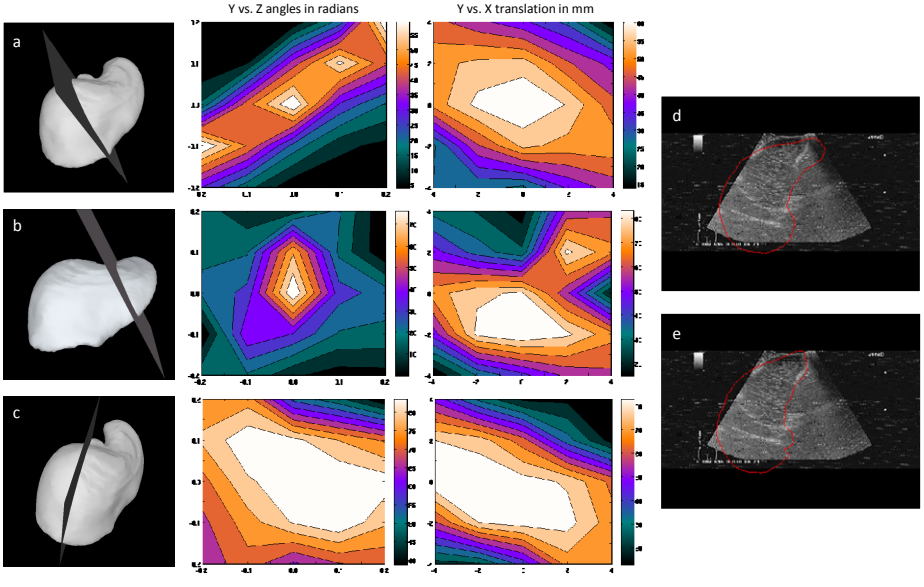


Fig. 2. (a, b, c) The Instantiation Index for the scan plane at three locations with respect to the silicone liver model. Optimisation for respiratory position 1 with the Aloka ultrasound image with (d) the initial contour and (e) the optimised contour with the first scan plane (a) as the optimised scan plane.

An examination of the Instantiability Index, with translation and rotation of the plane, for three different scan planes is shown in Fig. 2. The graphs show that a small change in translation of the plane is unlikely to affect the shape instantiation but a change in the orientation might, highlighting the need for scan plane optimisation. The third scan plane is the most forgiving to any perturbations but for this phantom, is inaccessible with the ultrasound probe. The result of the optimisation of a contour with an Aloka ultrasound image of the silicone phantom is also shown in Fig. 2. Fig. 3 demonstrates the results of the instantiation (mean error with standard deviation) when using the ultrasound data along the optimised scan plane versus the original scan plane. The errors are less when using the reinitialised shape instantiation method

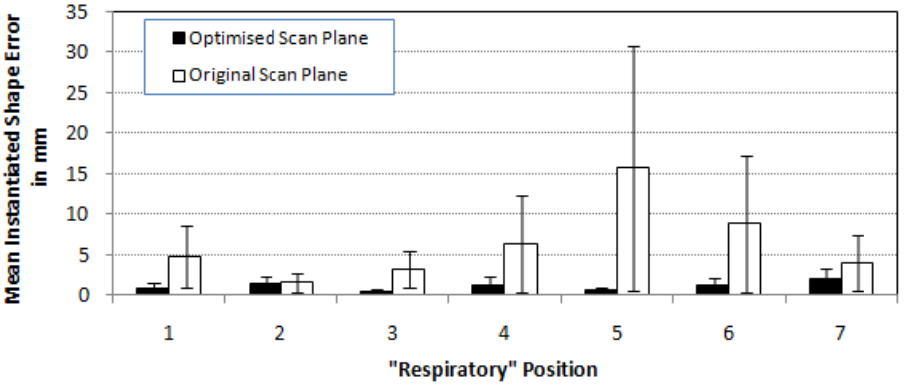


Fig. 3. The optimised instantiation mean errors, and standard deviations, and instantiation with the collected data at the original scan plane mean errors to the ground truth. The respiratory positions are numbered based on the number of lung inserts in the silicone model.

than when using the original scan plane. In the case of Position 5, the instantiation using the original scan plane failed, highlighting the necessity of optimising the scan plane intraoperatively.

3.2 Validation for Intraoperative Catheter Localisation

The setup for the IVUS experiment with the vascular silicone phantom can be seen in Fig. 4. For this second set of the IVUS data, the optimal location of the IVUS imaging probe was compared to that found in the ground truth CT images. A comparison of the graphs of the Instantiability Index, as the plane undergoes variation, at the descending aorta and near a bifurcation is shown in Fig. 5. A sharp peak is evident in the latter example, making that scan plane ideal for transducer localisation. Fig. 5 also shows the result of the optimisation of the model contours for Position 2 along with the final position of the imaging plane with respect to the geometry of the silicone vascular phantom. The error in the position of the IVUS catheter, calculated as the

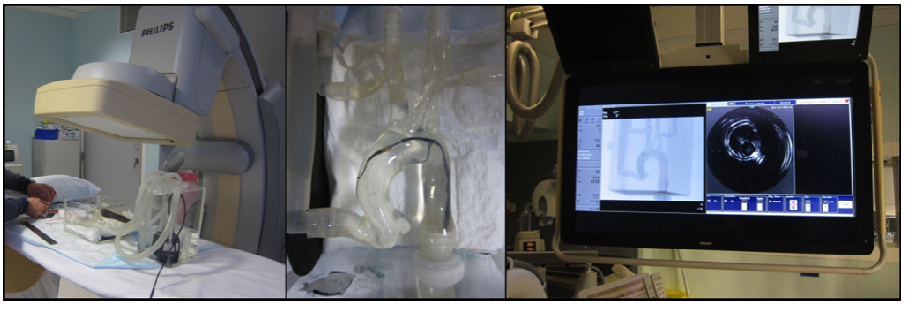


Fig. 4. The IVUS experiment setup. From left to right, the pulsatile silicone vascular phantom within the Philips scanner, the phantom with an IVUS catheter inserted, and the fluoroscopic image displayed with the IVUS image.

Euclidean distance between the calculated position and the ground truth found in the CT scan, across all 10 positions is shown in Table 1, along with their corresponding Instantiability Index Measures, with a higher measure indicating a graph with a greater peak. The errors were found to be generally less than 4 mm, with the exception of Position 4, when the IVUS catheter was located in one of the smaller branches of the phantom.

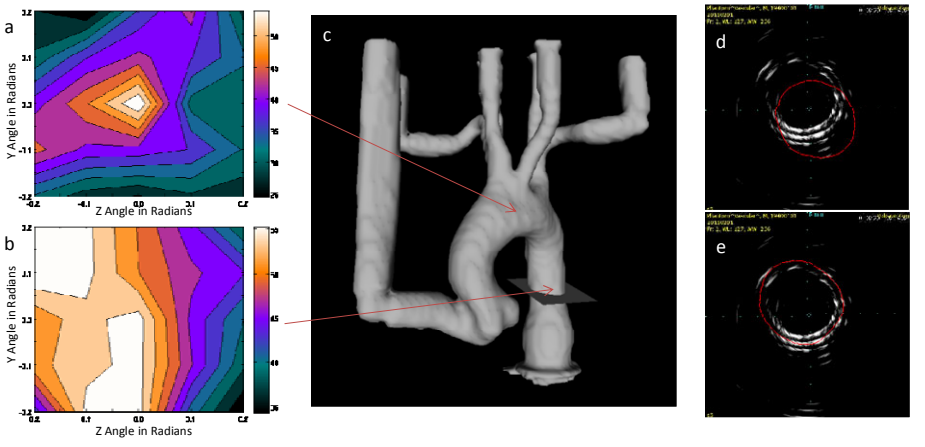
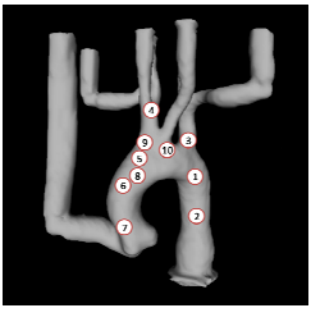


Fig. 5. The Instantiation Index (a) near a bifurcation of the artery and (b) at the descending aorta (where the lumen remains similar despite variations in the plane). Optimisation for Position 2 with the pre-processed IVUS image (d) with the initial contour overlaid, (e) the optimised contour, and (c) the optimised scan plane with the vascular phantom mesh.

Table 1. The errors in mm, along with the Instantiation Index Measure, of the estimated IVUS transducer position for the 10 different catheter positions in the silicone vascular phantom

Position	1	2	3	4	5
Error	3.05	2.32	2.58	6.86	3.20
IIM	0.97	1.19	1.08	1.10	1.52
Position	6	7	8	9	10
Error	2.01	1.83	3.70	4.08	1.14
IIM	1.48	1.27	1.23	1.07	1.49



With both silicone phantoms, there were many similar areas and hence many local minima where an optimisation method could be trapped. Simulated annealing was found to be the most successful optimisation technique for both datasets due to its ability to step out of local minima. However, it does not currently run in real-time.

4 Discussion and Conclusion

In this paper, we have proposed a new index for shape instantiation. We have demonstrated its use for both intraoperative anatomical reconstruction and probe tracking. The proposed instantiation index provides a measure of the amount of information in the scan plane to reduce the errors in instantiation and localisation. Local perturbations of the scan plane can be examined and a region of stable Indices is suitable for shape instantiation, with the use of an optimisation method to determine the exact plane captured with respect to the preoperative model. This optimised plane is used to reinitialise the predictive model, thus making prediction of the entire shape more accurate intraoperatively. A scan plane with a distinctive peak in the index, with respect to nearby scan planes, indicates a global optimum at that plane and hence is suitable for localisation within the preoperative model. The paper addresses a practical aspect of shape instantiation that was not considered in previous work and through the use of a novel shape instantiation index, improves both morphological reconstruction and device tracking.

Acknowledgments. The authors would like to thank Alessio Dore for his assistance in data collection.

References

1. Lee, S.-L., Chung, A., Lerotic, M., Hawkins, M.A., Tait, D., Yang, G.-Z.: Dynamic Shape Instantiation for Intra-operative Guidance. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6361, pp. 69–76. Springer, Heidelberg (2010)
2. Brij Koolwal, A., Barbagli, F., Carlson, C., Liang, D.: An Ultrasound-based Localization Algorithm for Catheter Ablation Guidance in the Left Atrium. The International Journal of Robotics Research 29, 643–665 (2010)
3. King, A.P., Rhode, K.S., Ma, Y., Yao, C., Jansen, C., Razavi, R., et al.: Registering Preprocedure Volumetric Images With Intraprocedure 3-D Ultrasound Using an Ultrasound Imaging Model. IEEE Transactions on Medical Imaging 29, 924–937 (2010)
4. Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P.: Numerical Recipes in C. Cambridge University Press, Cambridge (1992)
5. Mercier, L., Langø, T., Lindseth, F., Collins, D.L.: A review of calibration techniques for freehand 3-D ultrasound systems. Ultrasound in Medicine & Biology 31, 449–471 (2005)
6. Gegenfurtner, K.: PRAXIS: Brent's algorithm for function minimization. Behavior Research Methods 24, 560–564 (1992)
7. Prager, R.W., Rohling, R.N., Gee, A.H., Berman, L.: Rapid Calibration for 3-D Freehand Ultrasound. Ultrasound in Medicine and Biology 24, 855–869 (1998)
8. Lerotic, M., Lee, S.-L.: A Multimodal Silicone Phantom for Robotic Surgical Training and Simulation. In: The Hamlyn Symposium on Medical Robotics, London, UK, pp. 65–66 (2010)

Automatic Tracking of an Organ Section with an Ultrasound Probe: Compensation of Respiratory Motion

Caroline Nadeau[1], Alexandre Krupa[2], and Jacques Gangloff[3]

[1] Université de Rennes I, IRISA, 35042 Rennes, France
[2] INRIA Rennes-Bretagne Atlantique, IRISA, 35042 Rennes, France
[3] LSIIT, UMR 7005 CNRS-Université de Strasbourg I, 67400 Illkirch, France

Abstract. In minimally invasive surgery or needle insertion procedures, the ultrasound imaging can easily and safely be used to visualize the target to reach. However the manual stabilization of the view of this target, which undergoes the physiological motions of the patient, can be a challenge for the surgeon. In this paper, we propose to perform this stabilization with a robotic arm equipped with a 2D ultrasound probe. The six degrees of freedom of the probe are controlled by an image-based approach, where we choose as visual feedback the image intensity. The accuracy of the control law is ensured by the consideration of the periodicity of the physiological motions in a predictive controller. Tracking tasks performed on a realistic abdominal phantom validate the proposed approach and its robustness to deformation is assessed on a gelatin-made deformable phantom.

Keywords: visual servoing, ultrasound, motion compensation.

1 Introduction

Among the different medical imaging modalities, ultrasound (US) imaging is particularly attractive for providing intra-operatively real-time images. Indeed, this modality is safe and non-invasive for the patient and the US transducer, cheap and not cumbersome, can be easily used in an operating room. In particular, the US modality can provide to the surgeon a view of an organ or a tumor to reach throughout a minimally invasive surgery or a needle insertion procedure. In such applications, image-guided robotic systems could assist the surgeon insofar as they can follow an organ that undergoes physiological motions and thus automatically stabilize the US image of the target.

Some previous works dealing with US image-based robotic systems have focused on target tracking applications. In [1], the three in-plane motions of an US probe are controlled to automatically center in the US image the cross-section of an artery while an out-of-plane translation of the probe is manually performed. Five features extraction methods based on image similarity measures and contour segmentations are compared to track the artery center. In another work [2], the three translations of a XYZ stage robot equipped with two US probes and a HIFU transducer are controlled to follow a kidney stone while compensating for physiological motions during a lithotripsy procedure. In [3], an approach based on the speckle correlation observed in successive US

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 57–64, 2011.

images is detailed. However, a region of fully developped speckle has to be segmented and a step of learning of the speckle decorrelation curves is required. Moreover in these different approaches, the visual error minimization is realized with a proportional control law which does not benefit from the model of the disturbance.

On the other hand, physiological motions compensation using predictive controller in the control loop of the robotic system is a promising approach to take advantage of the repetitiveness of these disturbances. It has been developed in previous works where the tracking of natural or artificial landmarks is performed using an endoscopic camera. This visual feedback gives then a measure of the organ motion which is combined with a model of the periodic or quasi-periodic disturbance to predict and anticipate this motion. Many contributions focus on cardiac motion compensation using endoscopic cameras. In [4], Nakamura *et al.* present a tele-operated system to assist surgeons during beating heart surgeries. The slave arm is synchronized with the heart beats thanks to the visual feedback of a high speed camera and provides to the surgeon a stabilized image of the heart. In the same context of beating heart surgery, a motion prediction scheme is developed in [5] to increase the robustness of the detection and the tracking of natural landmarks on the heart surface in a laparoscopic view. More generally, Bebek *et al.* [6] describe improvements in motion canceling by taking into account biological signals (ECG) in the predictive algorithm. Recently, a comparison between various predictive filtering methods has been proposed in [7] to predict the motion of the mitral valve annulus. Under the assumption of a major 1D translational motion of this annulus, the motion compensation of one degree of freedom (dof) based on an extended Kalman filter is validated in situation of high noise, time delay and heart rate variability.

In this paper, we propose to control the six dof of a robotic arm holding a 2D US probe in order to follow a desired abdominal cross section while compensating for the respiratory motion. The considered applications are for instance the assistance for diagnoses or hepatic tumor biopsies where the liver and the tumors mainly undergo the respiratory motion [8]. Other clinic applications, such as prostate cancer brachytherapy have been identified in [3] that could benefit from such robotic image stabilization. To deal with the low quality of the US images, we propose to directly use the intensity of the image as visual feature, which has been successfully applied in camera-based visual servoing [10]. The structure of our paper is as follows: the second section details the principle of the intensity-based approach and the computation of the interaction matrix that links the intensity features variation to the motion of the 2D probe. The section 3 presents the proportional and predictive control laws implemented and compared with simulation results. The results of the tracking task performed on an abdominal and on a deformable phantom are presented in section 4 and included in a video[1].

2 US Image-Based Approach

The US images of the human body present a low quality due to the noise, called speckle, generated by the propagation of the US waves in the soft tissues. Because of this speckle, the processing of the US images is often more complex than with other

[1] `http://www.irisa.fr/lagadic/demo/demo-us-servoing-intensity/intensity_us_servoing.html`

imaging modalities, making the extraction of natural landmarks particularly complicated. We propose therefore to consider as visual features $\mathbf{s}$ the intensity values of the pixels of a region of interest (ROI) of the US image:

$$\mathbf{s} = \{I(1,1),...,I(u,v),...,I(M,N)\}, \tag{1}$$

where M and N are respectively the width and the height of the ROI and where $I(u,v)$ represents the intensity of the pixel of coordinates (u,v) in the US image.

Given $\mathscr{R}_p(\mathbf{x_p},\mathbf{y_p},\mathbf{z_p})$ the frame attached to the US probe where $(\mathbf{x_p},\mathbf{y_p})$ defines the image plane and $\mathbf{z_p}$ corresponds to the elevation axis, the coordinates $\mathbf{x} = (x,y,z)$ of the image features in this frame are such as:

$$(x,y,z) = (s_x(u-u_0), s_y(v-v_0), 0),$$

with (s_x,s_y) the image pixel size and (u_0,v_0) the pixel coordinates of the image center, and $z = 0$ since the considered visual features belong to the US image plane.

The computation of the interaction matrix $\mathbf{L_s}$ that links the variation of these image features to the motion of the probe is based on the constancy of the US wave reflexion by a physical 3D point. Given such a 3D point at the position $\mathbf{x}$ at the time t, which moves to the new position $\mathbf{x}+\mathbf{dx}$ at the time $t+dt$ and considering that the US reflexion is converted in an intensity value in a B-mode US image, the US reflexion conservation yields to the following intensity conservation equation:

$$I(\mathbf{x}+\mathbf{dx}, t+dt) - I(\mathbf{x},t) = 0. \tag{2}$$

We can expand this equation in the form of a first order Taylor series:

$$\frac{\partial I}{\partial x}dx + \frac{\partial I}{\partial y}dy + \frac{\partial I}{\partial z}dz + \frac{\partial I}{\partial t}dt = 0. \tag{3}$$

Then, the time variation of each pixel intensity $I(u,v)$ can be expressed as a function of the corresponding 3D point motion :

$$\dot{I} = -\nabla I \, \dot{\mathbf{x}}, \tag{4}$$

with $\nabla I = (\nabla I_x \;\; \nabla I_y \;\; \nabla I_z)$ the 3D image gradient, which is computed from the current probe image and at least two out-of-plane additional images. In practice, with a conventional 2D US probe mounted on a robotic arm, a small back and forth translational motion along the elevation direction is applied to acquire these additional images.

According to the kinematics fundamental relationship, the velocity of the 3D point $\dot{\mathbf{x}}$ is linked to the probe velocity $\mathbf{v}_c$ through the interaction matrix $L_{\mathbf{x}}$:

$$\dot{\mathbf{x}} = L_{\mathbf{x}}\,\mathbf{v}_c, \quad L_{\mathbf{x}} = \begin{bmatrix} -1 & 0 & 0 & 0 & -z & y \\ 0 & -1 & 0 & z & 0 & -x \\ 0 & 0 & -1 & -y & x & 0 \end{bmatrix}. \tag{5}$$

The interaction matrix L_I of size 1×6 associated to each visual feature is then defined as $L_I = -\nabla I \, L_{\mathbf{x}}$ and the complete interaction matrix $\mathbf{L_s}$, defined as $\dot{\mathbf{s}} = \mathbf{L_s}\mathbf{v}_c$, is built by stacking the $M \times N$ matrices L_I associated to each pixel.

3 Control Strategy

3.1 Classical Control Law

An image-based visual servoing control scheme consists in minimizing the error between a current set of visual features $\mathbf{s}$ and a desired one $\mathbf{s}^*$. In the case of a number p of visual features exceeding the number m of dof of the controlled system, a combination matrix $\mathbf{C}$ of size $m \times p$ and full rank m is introduced to define the task function to minimize $\mathbf{e}(t) = \mathbf{C}(\mathbf{s}(t) - \mathbf{s}^*)$.

In an eye-in-hand system, the instantaneous velocity applied to the imaging sensor $\mathbf{v}_c$ is computed from the task function and the interaction matrix $\mathbf{L_s}$. More particularly, the classical visual servoing control law described in [9] is based on an exponential decrease of the task function ($\dot{\mathbf{e}} = -\lambda\,\mathbf{e}, \lambda > 0$) and presents the best behavior with $\mathbf{C} = \widehat{\mathbf{L_s}}^+$, where $\widehat{\mathbf{L_s}}^+$ is the pseudo-inverse of the matrix estimate $\widehat{\mathbf{L_s}}$, which yields to:

$$\mathbf{v}_c = -\lambda \left(\widehat{\mathbf{L_s}}^+ \widehat{\mathbf{L_s}}\right)^{-1} \left(\widehat{\mathbf{L_s}}^+ (\mathbf{s}(t) - \mathbf{s}^*)\right) = -\lambda\, \widehat{\mathbf{L_s}}^+ (\mathbf{s}(t) - \mathbf{s}^*) \tag{6}$$

3.2 Predictive Controller

In the previous formulation, the controller does not take advantage of some knowledge on the disturbance model and more particularly of its periodicity. We propose then to implement in the control loop a predictive controller, based on the principle of the unconstrained Generalized Predictive Controller (GPC) [11], in order to anticipate the effect of the disturbance. We consider a system with multiple inputs $\mathbf{u}$ and outputs $\mathbf{y}$. The former are the six velocity components applied to the probe and the latter correspond to the image features observed in the current image $\mathbf{s}$. In practice, in order to avoid an excessive number of visual components and to ensure the controllability of the system, we introduce the combination matrix in the system loop to consider an output vector of dimension 6: $\mathbf{y} = \mathbf{Cs} = \widehat{\mathbf{L_s}}^+ \mathbf{s}$. The system is described with an ARIMAX model:

$$\mathbf{A}(z^{-1})\mathbf{y}(k) = \mathbf{B}(z^{-1})\mathbf{u}(k-1) + \frac{\mathbf{P}(z^{-1})}{\Delta(z^{-1})}\mathbf{b}(k), \tag{7}$$

where the polynomials $\mathbf{A}$ and $\mathbf{B}$ represent the theoretical model of the robotic system, approximated by a pure integrator. The noise is modeled by a white noise term $\mathbf{b}(k)$ which is colored by the polynomial matrix $\mathbf{P}$ and made non stationary by the operator $\Delta(z^{-1})$. In the classical GPC, this operator is taken as an integrator:

$$\Delta(z^{-1}) = 1 - z^{-1}.$$

In the case of periodic disturbances with a known fixed period, Gangloff et al [12] proposed to consider a repetitive noise model by modifying this operator as follows:

$$\Delta_R(z^{-1}) = 1 - \alpha z^{-T}, \;\; 0 < \alpha \leq 1$$

where α is a forgetting factor making the noise model repetitive with a period T corresponding to the number of sampling periods in one period of the disturbance signal.

To ensure that this repetitive noise model only affects the disturbance rejection and not the response to the visual reference, the new controller called R-GPC is a combination of two GPC (see Fig. 1). The controller GPC_1 is applied to the theoretical model of the system with no noise term which allows the robot to anticipate the future references $\mathbf{r} = \mathbf{C}\mathbf{s}^*$. The second controller GPC_2 includes the repetitive noise model previously defined to cancel the disturbance effect on the output vector.

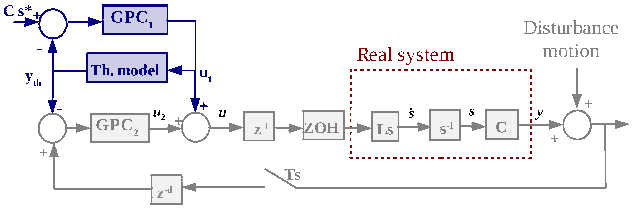


Fig. 1. Bloc diagram of the R-GPC including a control delay of one sampling period and a delay of d sampling periods due to the measure

The cost function minimized by the R-GPC approach is therefore composed of the cost functions of GPC_1 and GPC_2 as detailed in [12]:

$$J(\mathbf{u}=\mathbf{u_1}+\mathbf{u_2},k) = \sum_{j=N_1}^{N_2} \| \hat{\mathbf{y}}_{th}(k+j) - \mathbf{r}(k+j) \|^2 + \sum_{j=N_1}^{N_2} \| \hat{\varepsilon}(k+j) \|^2 + \lambda \sum_{j=1}^{N_u} \| \delta\mathbf{u_1}(k+j-1) \|^2 + \mu \sum_{j=1}^{N_u} \| \delta\mathbf{u_2}(k+j-1) \|^2 \quad (8)$$

where $N1, N2$ are the bounds of the prediction horizon and N_u is the length of the control horizon. λ and μ weight the control energies respectively for reference tracking and disturbance rejection. With this structure, the control signal $\mathbf{u}(t)$ applied to the real system is composed of the two components $\mathbf{u_1}(t)$ and $\mathbf{u_2}(t)$. The former corresponds to the input of the theoretical system model without disturbance that generates the theoretical output $\mathbf{y}_{th}$, the latter allows the rejection of the error $\varepsilon(t)$ due to noises and disturbance.

4 Results and Discussion

4.1 Simulation Results with a Human Liver

To validate our approach, we use a software simulator that we have developed to reconstruct and display a dense volume from a set of parallel images. In addition to this display functionality, the simulator allows us to move the reconstructed volume wrt a fixed Cartesian frame and to control a 2D virtual probe which generates an US image by cubic interpolation process. For the simulation experiments, an US complete volume of a human liver is loaded in the simulator. This volume is created from a set of 218 parallel images of resolution 308×278, the voxel size is $0.5 \times 0.5 \times 0.5 mm^3$.

In abdominal surgery, the hepatic motions are mainly induced by the respiration [8]. The cranio-caudal translation is the most important motion with an established range of 10 to 26*mm* and additional translations of about 8*mm* in antero-posterior and lateral directions are observed. In line with these data, a periodic translational motion is applied to the liver volume with amplitudes of 16*mm*, 5.5*mm* and 7*mm* along the x, y and z axes.

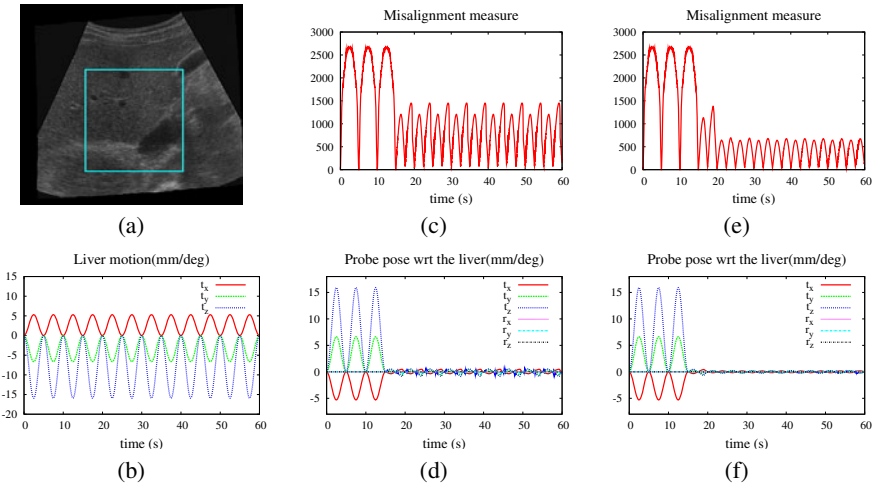


Fig. 2. Results of the tracking of a liver cross-section (a) that undergoes a periodic translational motion (b). For the same task, the results obtained with the proportional control law (c,d) and with the predictive one (e,f) are compared in term of visual error (c,e) and pose error (d,f).

The results of the simulation are displayed in Fig. 2. During three breathing periods no compensation is applied, then at $t = 15s$, the predictive controller R-GPC is launched and compared with the classical control law (6). From the first period the probe follows the liver motion and the misalignment measure defined as $\mathscr{C}(\mathbf{s}) = (\mathbf{s}-\mathbf{s}^*)^\top (\mathbf{s}-\mathbf{s}^*)$ is significantly reduced with both approaches (see Fig. 2(c) and (e)). With the R-GPC, the periodic disturbance is predicted and the accuracy of the tracking is improved.

4.2 Experimental Results with the Predictive Controller

The experiments have been performed with a robotic arm equipped with a convex 2-5 MHz US transducer on an abdominal phantom and a deformable one (see Fig. 3).

For the considered medical applications, we combine the visual control with a force control insofar as the US probe relies on the patient skin. We therefore add to the visual task a force constraint which guarantees a constant force of 1N applied to the phantom. The force control is used to servo the translational motion along the y-axis of the probe frame while the five remaining dof are controlled by the visual servoing control scheme. The Fig. 4 shows the results of one tracking experiment realized with the predictive controller where a sinusoidal perturbation is applied to the abdominal phantom with a period of 6s, mainly along the x and z axes in translation and around the y-axis

Fig. 3. Experimental setup with the abdominal (left) and gelatin-made (right) phantoms

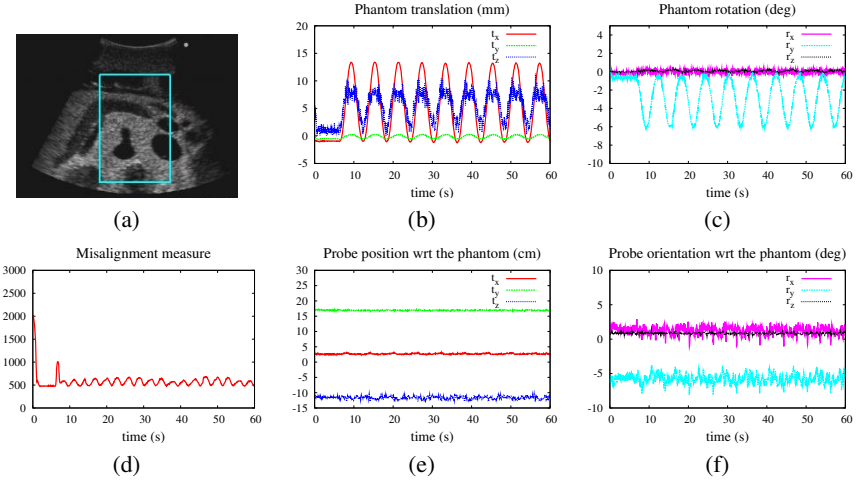
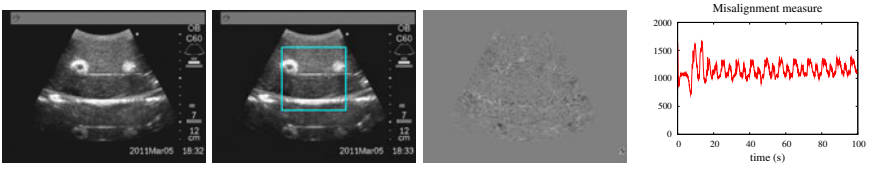


Fig. 4. Tracking of an abdominal cross section (a) while the phantom undergoes a periodic external motion applied at $t = 6s$ in translation (b) and rotation (c). The low value of the visual error (d) and the constant position (e) and orientation (f) of the probe wrt the phantom validate the task.

in rotation. The visual error is reduced from the second period, which guarantees the stabilization of the organ cross section in the US image during the disturbance cycles.

The robustness of the control law when dealing with deformable objects is assessed using a gelatin-made phantom filled with cereal fibers to create US speckle and olives modeling hepatic tumors (see Fig. 5). At $t = 5s$, a periodic translational motion of about $2cm$ is applied to this phantom which is also pressed against a wall to create local deformations. In this particular case, the relative pose of the probe wrt the phantom can not been measured and the success of the task is visually validated by the display of the view of the US probe and the corresponding image difference with the desired image (see Fig. 5(b) and (c)) at $t = 37s$, which corresponds to the maximum error in the misalignment measure (see Fig. 5(d)) observed during the tracking task.

The tracking results validate the intensity-based approach to deal with the stabilization of a 2D US image by compensating periodic disturbances with a predictive controller. The limitation due to the low quality of the US images is overcome by the choice



(a) (b) (c) (d)

Fig. 5. Tracking results with the deformable phantom of an US slice containing two tumors

of the image intensity as visual feature and the six dof of the probe are controlled thanks to the modeling of the interaction matrix associated to these features.

Acknowledgment. The authors acknowledge the support of the ANR project US-Comp of the French National Research Agency.

References

1. Abolmaesumi, P., Salcudean, S., Zhu, W., Sirouspour, M., DiMaio, S.: Image-guided control of a robot for medical ultrasound. IEEE Trans. on Rob. 18, 11–23 (2002)
2. Lee, D., Koizumi, N., Ota, K., Yoshizawa, S., Ito, A., Kaneko, Y., Matsumoto, Y., Mitsuishi, M.: Ultrasound-based visual servoing system for lithotripsy. In: IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, pp. 877–882 (2007)
3. Krupa, A., Fichtinger, G., Hager, G.D.: Real time motion stabilization with B-mode ultrasound using image speckle information and visual servoing. Int. J. of Rob. Res. 28, 1334–1354 (2009)
4. Nakamura, Y., Kishi, K., Kawakami, H.: Heartbeat synchronization for robotic cardiac surgery. In: IEEE Int. Conf. on Robotics and Automation, ICRA 2001, pp. 2014–2019 (2001)
5. Ortmaier, T., Groger, M., Boehm, D., Falk, V., Hirzinger, G.: Motion estimation in beating heart surgery. IEEE Trans. on Biomedical Engineering 52, 1729–1740 (2005)
6. Bebek, O., Cavusoglu, M.C.: Intelligent control algorithms for robotic-assisted beating heart surgery. IEEE Trans. on Rob. 23, 468–480 (2007)
7. Yuen, S.G., Kettler, D.T., Novotny, P.M., Plowes, R.D., Howe, R.D.: Robotic motion compensation for beating heart intracardiac surgery. Int. J. of Rob. Res. 28, 1355–1372 (2009)
8. Clifford, M.A., Banovac, F., Levy, E., Cleary, K.: Assessment of hepatic motion secondary to respiration for computer assisted interventions. Computer Aided Surgery 7, 291–299 (2002)
9. Espiau, B., Chaumette, F., Rives, P.: A new approach to visual servoing in robotics. IEEE Trans. on Rob. 8, 313–326 (1992)
10. Collewet, C., Marchand, E.: Photometric visual servoing. IEEE Trans. on Rob. 27 (2011)
11. Clarke, D.W., Mohtadi, C., Tuffs, P.S.: Generalized predictive control - Part I: The basic algorithm. Automatica 23, 137–160 (1987)
12. Gangloff, J., Ginhoux, R., De Mathelin, M., Soler, L., Marescaux, J.: Model predictive control for compensation of cyclic organ motions in teleoperated laparoscopic surgery. IEEE Trans. on Control System Technology 14, 235–246 (2006)

3D Ocular Ultrasound Using Gaze Tracking on the Contralateral Eye: A Feasibility Study

Narges Afsham, Mohammad Najafi, Purang Abolmaesumi, and Robert Rohling

University of British Columbia, Vancouver, BC, Canada
rohling@ece.ubc.ca

Abstract. A gaze-deviated examination of the eye with a 2D ultrasound transducer is a common and informative ophthalmic test; however, the complex task of the pose estimation of the ultrasound images relative to the eye affects 3D interpretation. To tackle this challenge, a novel system for 3D image reconstruction based on gaze tracking of the contralateral eye has been proposed. The gaze fixates on several target points and, for each fixation, the pose of the examined eye is inferred from the gaze tracking. A single camera system has been developed for pose estimation combined with subject-specific parameter identification. The ultrasound images are then transformed to the coordinate system of the examined eye to create a 3D volume. Accuracy of the proposed gaze tracking system and the pose estimation of the eye have been validated in a set of experiments. Overall system error, including pose estimation and calibration, are 3.12 mm and 4.68˚.

1 Introduction

Ultrasound has become an indispensible diagnostic tool for many ocular and orbital diseases [1-3]. Many studies show the ability of ultrasound to perform a pathological evaluation of the eyeball in its posterior segment [4, 5]. For example, in vitreo-retinal surgery, preoperative planning with ultrasound can be used [5]. Recent studies have also shown the benefit of ocular ultrasound in emergency medicine, such as diagnosing retinal detachment [6, 7]. There are some pathological characteristics that differentiate retinal detachment from other abnormalities such as vitreous hemorrhages [7]. Clinical manifestation of retina and vitreous detachments varies depending on where the adhesion is strongest [8]. These studies suggest that diagnosis, guidance and treatment would all benefit from imaging the entire posterior segment instead of a small portion.

Conventional ophthalmologic 2D ultrasound transducers have a small footprint and can image a portion of the posterior segment of the eye through the sclera. Since ultrasound has poor penetration of the cornea and lens, images are taken through the sclera when the gaze is deviated toward different positions [1]. However, the evaluation of pathological features using such 2D ultrasound makes the examination more complex and may contribute to false diagnoses [7]. In order to obtain views of the entire posterior segment, multiple ultrasound images must be captured for a variety of gaze directions. Moreover, the use of different gazes and the spherical geometry of the eyeball make it difficult to interpret the pose of each 2D ultrasound

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 65–72, 2011.

relative to the eye's coordinate system. The complexity of such examinations comes from the fact that the 3D pose of the eye should be known for each image acquisition [2]. In practice, the examiner has to visualize the intersection of the ultrasound plane with a mental 3D model of the eye and guess the eye pose to estimate where the absolute position of the image is located in the coordinate system of the eye [2]. There is a need for a new solution for this problem, especially since the eyelid is closed during the examination, so the eye pose is difficult to discern visually.

It has been suggested that a dedicated 3D ultrasound transducer facilitates the assessment of certain posterior ocular abnormalities [9]. However, even in the case of a specialized 3D ultrasound transducer, only a portion of the posterior surface would be imaged from each pose of the eye and it does not eliminate the localization problem associated with a gaze-deviated examination. Moreover, this challenge cannot be solved by tracking the ultrasound transducer relative to a fixed base because the eye changes its pose during the examination.

We propose a novel system using a combination of a camera, an ultrasound transducer, and a subject-specific eye pose estimation method to tackle this problem. The pose of the examined eye is estimated from the pose estimation of the contralateral eye during the imaging process. Eye parameters for each subject and in different lightening conditions are modeled together with pose estimation in an attempt to achieve high localization accuracy. The method of model-based gaze tracking proposed in [10] is adopted and modified to replace the parameters based on population averages, such as cornea radius and pupil radius, with parameters specific to the subject. The proposed system aims to produce accurate 3D ultrasound reconstruction of the posterior segment of the eye from a set of 2D images acquired from different eye gazes.

2 Methods and Materials

2.1 Overview

As mentioned, the goal is to find the transformation from the ultrasound image to the examined eye's coordinate system. This transform, ${}^{E2}\boldsymbol{T}_{\boldsymbol{US}}$, can be defined as:

$$ {}^{E2}\boldsymbol{T}_{\boldsymbol{US}} = {}^{E2}\boldsymbol{T}_{\boldsymbol{E1}}\,{}^{E1}\boldsymbol{T}_{\boldsymbol{C}}\,{}^{C}\boldsymbol{T}_{\boldsymbol{US}} \tag{1} $$

where ${}^{C}\boldsymbol{T}_{\boldsymbol{US}}$, ${}^{E1}\boldsymbol{T}_{\boldsymbol{C}}$ and ${}^{E2}\boldsymbol{T}_{\boldsymbol{E1}}$ are the transformations from ultrasound image to camera, camera to Eye1, and Eye1 to Eye2, respectively (Fig. 1(a)). Eye2 is the examined eye and Eye1 is the contralateral eye. ${}^{C}\boldsymbol{T}_{\boldsymbol{US}}$ is a fixed transformation that is found by a separate calibration process. ${}^{E1}\boldsymbol{T}_{\boldsymbol{C}}$ is determined for each pose of Eye1 by gaze tracking together with estimating subject-specific eye parameters. These parameters are the cornea center, $\boldsymbol{c}$, cornea radius, r_c, pupil center, $\boldsymbol{p}$, pupil radius, r_p, and cornea-center-to-pupil-center distance, r_{cp}. The optical axis is a vector that joins $\boldsymbol{c}$ and $\boldsymbol{p}$, and the visual axis passes through $\boldsymbol{c}$ and hits the fovea (Fig. 2(a)). To find ${}^{E2}\boldsymbol{T}_{\boldsymbol{E1}}$ some parameters should first be calibrated in an experiment with two open eyes for each individual. One is the transform from the rest pose of Eye1 to Eye2, ${}^{Er2}\boldsymbol{T}_{\boldsymbol{Er1}}$. The rest pose is defined as the eye pose when it is looking at a distant target and head is upright and static [11]. The other is the angular offset between the visual axis and the optical axis, β_v [12] (Fig. 2(a)). The visual axes of the two eyes intersect at

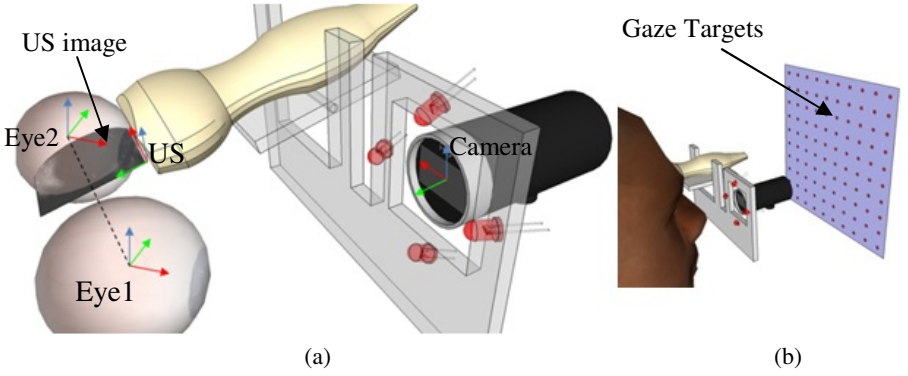


Fig. 1. System overview. (a) Each US image should be transformed into the coordinate system of Eye2. The pose of Eye1 is tracked in the coordinate system of the camera using glint tracking from four LEDs. The US to camera transform is fixed during the experiment. The transform from Eye1 to Eye2 is determined by a subject-specific model. (b) A subject fixates at several target points to create a wide range of eye movements during an examination.

the fixation points. Considering this fact for each target point, and having ${}^{Er2}\boldsymbol{T}_{Er1}$ and the angular offset for both eyes, β_{v1}, β_{v2}, the Eye1 to Eye2 transformation, ${}^{E2}\boldsymbol{T}_{E1}$, can be determined. The detailed calculations of these stages are described in the following sections. Given space limitations only the key equations are described in full.

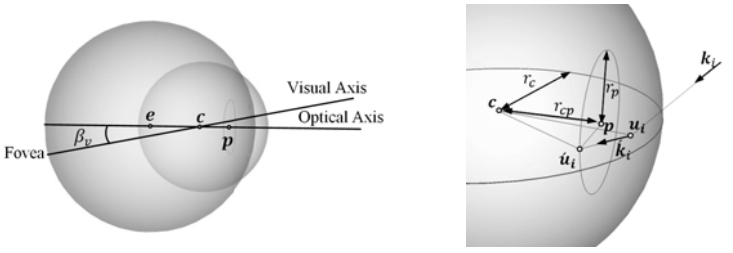
2.2 Calibration

The subject is asked to look at a distant point with both eyes open. The rest position of both eyes is found from gaze tracking of the two eyes with a single camera placed farther from the face and ${}^{Er2}\boldsymbol{T}_{Er1}$ is measured using gaze tracking on both eyes as a subject is guided to a distant target. Then the subject is asked to look at several known target points on a board and the optical axis is found for each target (Fig. 1(b)). Then the method as described in [13] is used to calculate β_{v1} and β_{v2}.

For ultrasound to camera calibration, ${}^{C}\boldsymbol{T}_{US}$, the single-wall calibration method [14] is used with the modification of using a checker-board pattern on the portion of the wall surface outside the water bath so that it can be seen by the camera. The camera's intrinsic parameters are calibrated with the *Camera Calibration Toolbox* [15], which uses images of a checker-board pattern.

2.3 Combined Pose Estimation and Eye Model Parameter Identification

The primary application of gaze tracking is in the field of human-computer interfaces. In the video-based gaze tracking literature, those that give the full 3D pose of the eye, instead of simply a gaze intersection with a monitor, are of interest here. We have implemented a gaze tracking system based on a single camera and multiple IR light sources (LEDs) [10]. In this method, the glints are used to estimate the position of cornea center. The circumference of the pupil in the image is refracted into the eye according to Snell's law. With a known position of the cornea, it is possible to



(a) (b)

Fig. 2. (a) 3D model of the eye. $\boldsymbol{e}$, $\boldsymbol{c}$, and $\boldsymbol{p}$ are eye, cornea, and pupil centers (b) The refraction at the cornea surface is shown. $\boldsymbol{k_i}$ is the incident ray and $\boldsymbol{k_i'}$ is the refracted ray.

calculate the center of the pupil. The calculations of the cornea center and pupil center are the same as explained in [10] with some modifications to make it subject-specific by including identification of r_p, r_c and r_{cp} in the calculations.

Geometric Eye Model. Fig. 2 (a) shows the geometry of the eye modeled as two spheres, the eyeball and the cornea [12]. As a result of a fixed head position, the center of the eyeball, $\boldsymbol{e}$, remains fixed during the experiments. This confines the transformation of the eye from one position to another as a pure rotation. By knowing the optical axis and applying Listing's law, it is possible to determine the full pose of the eye [11]. Listing's law states that the eye's axis of rotation from the rest pose to any arbitrary pose is restricted to a single head-fixed plane, known as Listing's plane [16]. To determine the optical axis for each pose, the cornea and pupil centers should be estimated. Finally ${}^{E1}\boldsymbol{T_C}$ and ${}^{E2}\boldsymbol{T_{E1}}$ are calculated based on the equations in [11].

Cornea Parameter Estimation. There exists an auxiliary coordinate system in which the $\boldsymbol{xz}$ plane includes the camera origin, $\boldsymbol{o}$, light source, $\boldsymbol{l_i}$, cornea center, $\boldsymbol{c}$, its corresponding glint in the image, $\boldsymbol{g_i}$, and its 3D position, $\boldsymbol{g_i'}$ [10] (see Fig. 3(b)). The subscript i ranges from 1 to 4 light sources. Each light source vector resides along the $\boldsymbol{x}$ direction of its auxiliary coordinate. The transform from the camera coordinate to the auxiliary coordinate is defined by a rotation matrix, $\boldsymbol{R_i}$, in terms of $\boldsymbol{l_i}$ and $\mathbf{c_i'}$ (cornea center in the auxiliary coordinate) such that $\boldsymbol{c} = \boldsymbol{R_i}\, \boldsymbol{c_i'}$.
$g_i'd_i$ is the bisector of the $og_i'l_i$ angle based on the reflection theorem. As a result of the bisector equation, one may write:

$$\frac{\sqrt{g_{ix}'^2+g_{iz}'^2}}{d_i} = \frac{\sqrt{(g_{ix}'-l_i)^2+g_{iz}'^2}}{l_i-d_i}\,, \quad \xrightarrow{yields} \quad d_i = \frac{l_i\sqrt{g_{ix}'^2+g_{iz}'^2}}{\sqrt{g_{ix}'^2+g_{iz}'^2}+\sqrt{(g_{ix}'-l_i)^2+g_{iz}'^2}}\,. \tag{2}$$

The triangulation for $d_ig_i'h_i$ and $g_i'c_i'f_i$ results in:

$$\frac{r_c}{r_c+\sqrt{(g_{ix}'-d_i)^2+g_{iz}'^2}} = \frac{c_{iz}'-g_{iz}'}{c_{iz}'}\,, \text{ and} \tag{3}$$

$$\frac{r_c}{r_c+\sqrt{(g_{ix}'-d_i)^2+g_{iz}'^2}} = \frac{c_{ix}'-g_{ix}'}{c_{ix}'-d_i}\,. \tag{4}$$

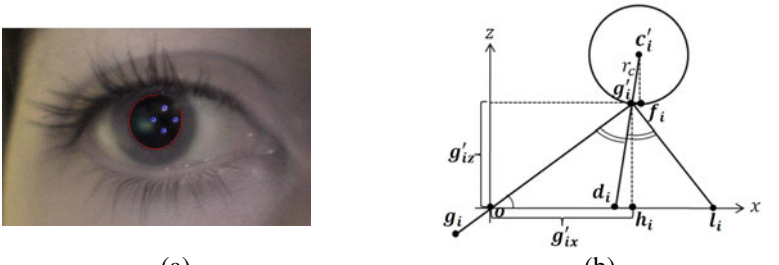
where r_c is the cornea radius. Since the angle between ol_i and og_i is known from the pinhole model of the camera, it gives another equation:

$$\frac{g'_{iz}}{g'_{ix}} = tan\,\alpha_i\,. \tag{5}$$

Eqs. (2) to (5) can be written for each light source and they result in $4n$ equations (n = number of light sources). Since the auxiliary coordinate system has the same origin as the camera, the transform from the camera coordinate system to the auxiliary system is defined by the rotation matrix, $\boldsymbol{R_i}$. Another set of equations is produced by the constraint that the transformation of auxiliary cornea centers to the camera coordinate system are all the same:

$$\boldsymbol{R_i^{-1}}\,\boldsymbol{c'_i} = \boldsymbol{R_i^{-1}}\,\boldsymbol{c'_j}\,. \tag{6}$$

In summary, by considering n light sources, the number ($4n + 3\binom{n}{2}$) of equations equals or exceeds the number ($5n + 1$) of unknowns (d_i, g'_{ix}, g'_{iz}, c'_{ix}, c'_{iz}, r_c). Although two light sources are sufficient to solve for the cornea center, we used four sources to get more accurate results [13].



(a) (b)

Fig. 3. (a) An example of camera image with four glints (blue) and extracted pupil perimeter (red) (b) Auxiliary coordinate system. $\boldsymbol{c'_i}$ is the auxiliary cornea center. $\boldsymbol{g'_i}$ is the glint of light source l_i in the image, and $\boldsymbol{g_i}$ is its 3D position.

Pupil Parameter Estimation. Fig. 2(b) depicts the geometry of the pupil center estimation problem. Instead of fitting an ellipse on the image of the pupil and estimating the pupil center from its refracted center in the image, we directly solve for pupil center as well as pupil radius, $\boldsymbol{r_p}$, and the distance between pupil-cornea centers, $\boldsymbol{r_{cp}}$. The direction of the vector from a point residing on the circumference of the pupil to the point of refraction on the surface of the cornea ($\boldsymbol{k_i}$) is known from the pinhole camera model, so the intersection of the ray with the cornea sphere surface ($\boldsymbol{u_i}$) is easily determined. $\boldsymbol{k'_i}$ is the direction of the refracted ray and it is known by considering Snell's law and the cornea's refractive index ($\boldsymbol{n_{cornea}}$ =1.336). The following gives three equations for each point on the pupil circumference, where $\boldsymbol{s_i}$ is an unknown parameter:

$$\boldsymbol{u'_i} = \boldsymbol{u_i} + s_i\boldsymbol{k'_i}. \tag{7}$$

$\boldsymbol{u_i'}$ is the intersection of the refracted ray and pupil perimeter , so we have:

$$\|\boldsymbol{u_i'} - \boldsymbol{p}\|^2 = r_p^2. \tag{8}$$

The constraint between the pupil perimeter and the cornea center gives:

$$\|\boldsymbol{u_i'} - \boldsymbol{c}\|^2 = r_{cp}^2 + r_p^2. \tag{9}$$

By adding the constraint of the distance between cornea center and pupil center:

$$\|\boldsymbol{c} - \boldsymbol{p}\|^2 = r_{cp}^2 . \tag{10}$$

By considering Eqs. (7) to (10) for at least four points ($n \geq 4$) on the pupil, the number ($5n + 1$) of equations equals or exceeds the number ($4n + 5$) of unknowns ($\boldsymbol{u_i'}$, $s_i, \boldsymbol{p}, r_p, r_{cp}$). Since the camera is close to the eye, the pupil perimeter points are easily labeled using Canny edge detection and if some portion of the pupil perimeter are missing in the image, a solution can still be found. In practice, at some far deviated gazes, the eyelid partly hides the perimeter of the pupil.

2.4 Experiments

The experimental setup consists of ultrasound imaging and gaze tracking components. The subject is positioned with a standard chin and forehead support to keep the head fixed related to the camera and the ultrasound transducer. The ultrasound system used is a Sonix MDP (Ultrasonix Medical Corporation, Richmond, BC, Canada) equipped with an L14-5/38 10 MHz linear 2D ultrasound transducer.

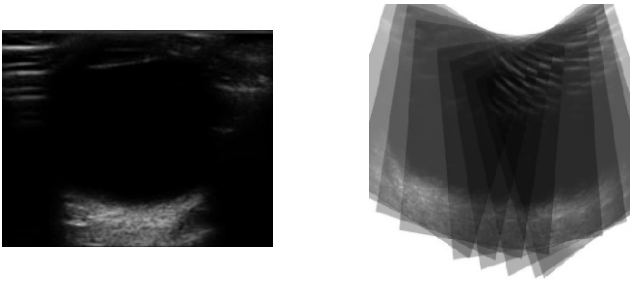
The ultrasound transducer is positioned on the eyelid using coupling gel, as standard practice. The transducer can be tilted around the elevation axis to obtain different imaging views. Then the transducer is fixed to the suitable position with respect to the chin rest. The gaze tracking part includes a high definition USB camera[1] (LifeCam Cinema, Microsoft Co., USA) and four infrared LEDs to make four bright glints in the image of the eye.

From Eq. 1, it is clear that the accuracy of the camera to ultrasound calibration, ${}^{C}\boldsymbol{T_{US}}$, can be easily separated from the gaze tracking results. To evaluate our overall system accuracy, we therefore give these as three separate components of the total error. In this way, the errors from our eye pose estimation ${}^{E2}\boldsymbol{T_{E_1}}$ and ${}^{E1}\boldsymbol{T_C}$ can be compared to a gold standard measurement of ${}^{E2}\boldsymbol{T_C}$ using eye tracking directly on the examined eye with the eyelid open. It also allows direct comparison of our novel combined modeling and tracking method with standard gaze tracking methods using parameters based on population averages.

Ultrasound to camera calibration is therefore performed first. Then in an experiment with two eyes open, the subject pursues target points on a board. The target points are evenly distributed (2 cm spacing) and the plane is 25 cm away from the camera. As the subject fixates at each target point, the gaze tracking system tracks the optical axis of the eyes and β_{v1} and β_{v2} are calibrated. Also, the rest pose of the eyes and ${}^{Er2}\boldsymbol{T_{Er1}}$ are determined using a distant gaze.

To evaluate ${}^{E2}\boldsymbol{T_{E1}}$ estimation, first it is directly calculated from the optical axis of both eyes. Then, it is estimated based on the contralateral eye from the fixation point

[1] The infra-red filter was removed.



(a) (b)

Fig. 4. (a) An example 2D ultrasound image of the posterior segment of the eye. (b) Ultrasound frames transformed in the coordinate system of the eye for a set (n=15) of eye gazes.

and the calibrated parameters. The fixation point is found from the intersection of Eye1's visual axis with the target plane. The unknown visual axis of Eye2 passes through its cornea center and the fixation point. The cornea center, the optical and the visual axes can be related to their rest pose with a single transform. This transform is found so that the visual axis passes through the target point. The difference in the eye orientation for $^{E2}\boldsymbol{T}_{E1}$ between our method and the gold standard are then calculated. The equations are solved with nonlinear least squares using the Levenberg-Marquardt method. As a feasibility study, one subject is examined and the errors are calculated (Table 1). Since the head is fixed, the eye center, $\boldsymbol{e}$, remains fixed. The standard deviation in estimation of $\boldsymbol{e}$ is also calculated as a measure of precision of the gaze tracking method.

3 Results

The residual errors (the minimized error) from the ultrasound to camera calibration are 0.2, 0.3, 0.5 mm, 0.2˚, 0.2˚, and 0.6˚. The accuracy of a single-wall calibration technique based on minimizing the residual errors is well documented [14]. Table.1 summarizes the differences in $^{E2}\boldsymbol{T}_{E1}$ from the gold standard. The standard deviation of the error in β_{v1} and β_{v2} estimations are 1.28˚ and 1.58˚, respectively. RMS error in eye center estimation for n=30 target points with our method is 2.91 mm compared to 3.72 mm error with model-based method using parameters based on population averages. Adding the errors of independent sources in Eq.1 gives an overall accuracy of 3.12 mm and 4.68˚.

By positioning each ultrasound image in the coordinates system of the eye for each pose, a 3D image of the eye is constructed. Fig. 4 demonstrates that the angular view of the posterior segment of the eye has been increased.

Table 1. The differences in $^{E2}T_{E1}$ from the gold standard in deg

Estimation errors of rotation parameters of $^{E2}T_{E1}$	$\boldsymbol{\alpha}$ **(Roll)**	$\boldsymbol{\beta}$ **(Pitch)**	$\boldsymbol{\gamma}$ **(Yaw)**
Mean	2.04	-0.32	-1.42
STD	3.15	1.91	1.61

4 Discussion and Conclusion

In this work a novel system has been proposed to determine the absolute pose of a 2D ultrasound image in the 3D coordinate system of the eye in a gaze-deviated eye examination. To this end, a subject-specific eye pose estimation method has been developed. The method is based on single camera eye gaze tracking.

The accuracy of the overall system is mainly governed by the accuracy of the camera to ultrasound calibration and the accuracy of eye pose estimation from the contralateral eye. Experimental results show errors 3.12 mm and 4.68°compared to the gold standard. The gold standard is eye gaze tracking directly on the examined eye, and itself has an error of approximately 1°[10]. The main sources of error are related to camera quality, position of the light sources and eye modeling.

Accurate 3D reconstruction of the eye's posterior segment and the assessment of dynamic ultrasound images of the eye are the other future discussion for this research.

Acknowledgments. Supported by the BC Innovation Council ICSD program.

References

1. DiBernardo, C.W., Schachat, A.P., Sharon, Fekrat, M.D.: Ophthalmic Ultrasound: A Diagnostic Atlas. Thieme Medical Publishers (1998).
2. Garcia Jr, J.P.S., Finger, P.T., Rosen, R.B.: Dynamic Ophthalmic Ultrasonography: A Video Atlas for Ophthalmologists and Imaging Technicians. Lipp. Williams & Wilkins, Baltimore (2009)
3. Fledelius, H.C.: Ultrasound in ophthalmology. Ultrasound Med. Biol. 23, 365–375 (1997)
4. Fisher, Y., Hanutsaha, P., Tong, S., Fenster, A., Mazarin, G., Mandava, N.: Three-dimensional ophthalmic contact B-scan ultrasonography of the posterior segment. Retina 18, 251–256 (1998)
5. Muldoon, T.O., Will, D.V., Garcia, J.P., Walsh, J.B., Rosen, R.B., Rosenthal, J.L., Gentile, R.C.: 3D B-mode Ultrasound of the Vitreous Prior to Vitrectomy Surgery. Invest. Ophthalmol. Vis. Sci. 44, 3625 (2003)
6. Shinar, Z., Chan, L., Orlinsky, M.: Use of Ocular Ultrasound for the Evaluation of Retinal Detachment. The Journal of Emergency Medicine 40, 53–57 (2011)
7. Yoonessi, R., Hussain, A., Jang, T.B.: Bedside Ocular Ultrasound for the Detection of Retinal Detachment in the Emergency Dept. Acad. Emerg. Med. 17, 913–917 (2010)
8. Sebag, J.: Anomalous Posterior Vitreous Detachment: A Unifying Concept in Vitreo-Retinal Disease. Graefes Arch. Clin. Exp. Ophthalmol. 242, 690–698 (2004)
9. Downey, D.B., Nicolle, D.A., Levin, M.F., Fenster, A.: Three-dimensional Ultrasound Imaging of the Eye. Eye (Lond) 10, 75–81 (1996)
10. Hennessey, C.: A Single Camera Eye-Gaze Tracking System with Free Head Motion. In: Proceedings of etra: Etra Symposium, pp. 87–94 (2006)
11. Horaud, R., Hansard, M.: Cyclorotation Models for Eyes and Cameras. IEEE Transaction on Systems Man and Sybernetics 40, 151–161
12. Villanueva, A., Daunys, G., Hansen, D.W., Böhme, M., Cabeza, R., Meyer, A., Barth, E.: A Geometric Approach to Remote Eye Tracking. Univ. Access Inf. Soc. 8, 241–257 (2009)
13. Villanueva, A., Cabeza, R.: Models for Gaze Tracking Systems. In: EURASIP, pp. 1–16 (2007)
14. Prager, R.W., Rohling, R.N., Gee, A.H., Berman, L.: Rapid Calibration for 3-D Freehand Ultrasound. Ultrasound in Med. & Bio. 24, 855–869 (1998)
15. Bouguet, J.-Y.: Visual Methods for Three-dimensional Modeling, Phd Thesis (1999)
16. Hepp, K.: Theoretical Explanations of Listing's law and their Implication for Binocular Vision. Vision Res. 35, 3237–3241 (1995)

Interactive 3D Visualization of a Single-View X-Ray Image

Matthias Wieczorek[1], André Aichert[1], Pascal Fallavollita[1], Oliver Kutter[2], Ahmad Ahmadi[1], Lejing Wang[1], and Nassir Navab[1]

[1] Chair for Computer Aided Medical Procedures (CAMP), Technische Universität München, Munich, Germany
[2] Siemens Corporate Research, Princeton, NJ, USA
wieczore@cs.tum.edu

Abstract. In this paper, we present an interactive X-Ray perceptual visualization technique (IXPV) to improve 3D perception in standard single-view X-Ray images. Based on *a priori* knowledge from CT data, we re-introduce lost depth information into the original single-view X-Ray image without jeopardizing information of the original X-Ray. We propose a novel approach that is suitable for correct fusion of intra-operative X-Ray and ultrasound, co-visualization of X-Ray and surgical tools, and for improving the 3D perception of standard radiographs. Phantom and animal cadaver datasets were used during experimentation to demonstrate the impact of our technique. Results from a questionnaire completed by 11 clinicians and computer scientists demonstrate the added value of introduced depth cues directly in an X-Ray image. In conclusion, we propose IXPV as a futuristic alternative to the standard radiographic image found in today's clinical setting.

1 Introduction

For almost all medical imaging modalities, current state of the art visualization systems provide advanced techniques for real-time *direct volume rendering* (DVR), which is used for diagnosis, planning, and intra-operative procedures [5]. Fused visualization of co-registered, multi-modal image data of the patient and traditional surface based rendering, e.g. tracked instruments, implant models, or segmentation has become common practice [4,11]. Furthermore, depth perception has been an important field of research in computer graphics. Especially for DVR many methods for enhancing depth cues have been proposed [1,2,4,11,12]. For translucent volumes (such as X-Ray) an existing method uses stereoscopy [8]. However, little interest has been taken in transferring the advances in DVR to introduce additional depth cues to single X-Ray images. The intensities displayed in an X-Ray image go back to the attenuation of a complete ray traversal through the body and do not correspond to a specific depth along the ray. This also makes fusion of rendered objects with X-Ray images more difficult.

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 73–80, 2011.

Current state of the art methods use either 2D blending [10] with a user defined blending factor (i.e. common 2D/3D registration algorithms) or depth-of-field based 2D blending, which defines the blending factor per pixel according to the depth of the geometry. Both methods have the same major drawback: they do not account for the physics of X-Ray. Thus, the resulting fused image will provide inaccurate and frequently misleading depth cues (see Fig. 3 (center), (right)) making it hard for the user to correctly estimate distances and positions of objects from the fused view. The objective of our work is to propose a novel visualization technique for improving visual perception in general and depth perception in particular in medical X-Ray images. Our idea termed *Interactive X-Ray Perceptual Visualization* (IXPV) allows the user to interactively manipulate a single-view X-Ray image by varying depth. This is achieved by retrieving *a priori* knowledge of absorptive properties from CT data (pre-operative or atlas). Through phantom and animal cadaver experimentation and user study, we show that the IXPV technique introduces depth cues in X-Ray image visualization to disambiguate the ordering of internal structures or instruments used in everyday diagnostic and intraoperative scenarios. Possible applications for this method range from X-Ray based medical Augmented Reality to fusion of X-Ray and DVR renderings of volumetric data, such as PET scans.

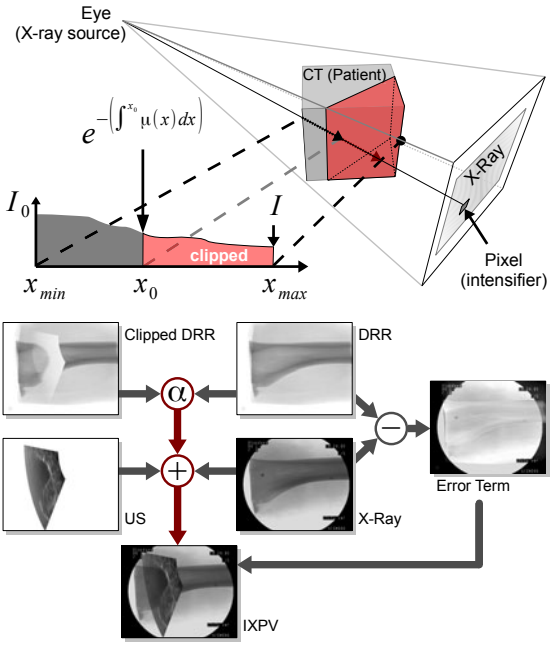


Fig. 1. (top) Classical analogy of DRR creation and X-Ray, after the CT (cube) has been registered. The figure also shows the effect of clipping the red portion of the CT on the attenuation function. (bottom) The major steps in IXPV rendering. α represents the coefficient used to modulate X-Ray and US (red arrows and +). Also shown is the error term from Section 2.7.

2 Methods

The proposed method aims at transferring information from co-registered CT data directly into X-Ray images. There exist several algorithms facing the problem of CT/X-Ray co-registration. There are marker, landmark and intensity

based [9,10] methods. One could even think of using an Atlas CT i.e utilizing methods as proposed in [3] or [7]. In our method, we assume that we have a registration between CT and X-Ray using any of the aforementioned methods. All derivations below shield one goal: calculation of the X-Ray intensities at any depth along a certain ray by modulating the X-Ray image and not simply copying pixel values from *Digitally Reconstructed Radiograph* (DRR).

2.1 Beer-Lambert Law and X-Ray Intensities at Specific Depth

For an emitted dose of X-Ray radiation I_0, the Beer-Lambert law for linear X-Ray attenuation describes the intensity I at the end of a ray:

$$I = I_0 * e^{-\int \mu(x)dx} \tag{1}$$

where $\mu : [0,1]^3 \rightarrow [0,1]$ is the X-Ray absorption function. A CT is essentially a discrete volume of absorption values measured in Hounsfield Units. Thus the absorption is given by $\mu_{\text{CT}} = g \circ f$, where $f : [0,1]^3 \rightarrow [0,1]$ is a function defined by the CT and $g : [0,1] \rightarrow [0,1]$ is the *transfer function* (TF) mapping the values from CT to the correct absorption value [9].

For a moment, assume X-Rays were cut off at some specific depth. We are interested in the X-Ray intensity up to a specific point x_0 along this ray, as if the intensifier were placed there (compare Fig. 1 (top)). As the absorption function is monotonic decreasing this is valid and with Eq. 1 we derive:

$$I = I_0 * e^{-(\int^{x_0} \mu(x)dx + \int_{x_0} \mu(x)dx} = I_0 * e^{-\int^{x_0} \mu(x)dx} * e^{-\int_{x_0} \mu(x)dx} \tag{2}$$

Since luminance values are $\in [0,1]$, choose $I_0 = 1$, for maximal radiation:

$$e^{-\int^{x_0} \mu(x)dx} = I / e^{-\int_{x_0} \mu(x)dx} \quad \text{and} \quad e^{-\int_{x_0} \mu(x)dx} = I / e^{-\int^{x_0} \mu(x)dx} \tag{3}$$

2.2 Transfer of Information to X-Ray

Assuming exact registration (for discussion on error see section 2.7), this leads to an approximative absorption function $\mu_{\text{CT}} \approx \mu \quad \forall x \in [0,1]^3$. Using Eq. 3:

$$e^{-\int^{x_0} \mu(x)dx} = I_{\text{X-Ray}} / e^{-\int_{x_0} \mu(x)dx} \approx I_{\text{X-Ray}} / \left(\frac{I_{\text{CT}}}{e^{-\int^{x_0} \mu_{\text{CT}}(x)dx}}\right) \tag{4}$$

$$\Rightarrow e^{-\int^{x_0} \mu(x)dx} \approx \underbrace{I_{\text{X-Ray}}}_{\text{X-Ray value}} * \underbrace{\frac{e^{-\int^{x_0} \mu_{\text{CT}}(x)dx}}{e^{-\int \mu_{\text{CT}}(x)dx}}}_{\text{blending factor}} \tag{5}$$

We derive a blending factor for the original X-Ray image, which we use to modulate intensity information of geometry or another modality for a specific depth. The blending factor is defined as the quotient of two DRR-passes. The full DRR (denominator) I_{CT} has to be computed once for a given X-Ray image and the clipped DRR (numerator) has to be updated if x_0 (i.e. the object boundary) changes and only for pixels where an object is present.

2.3 Interactive X-Ray Perceptual Visualization (IXPV)

Given the registration between CT or atlas CT and the X-Ray images and the method proposed in 2.2, we can interactively manipulate the depth information across the co-registered X-Ray image. In fact, in general terms many interactive manipulations done on particular 3D views of CT images could be duplicated using only modulated X-Ray values. The image could be a static X-Ray view or a running fluoroscopic sequence.

2.4 Clipping X-Ray

The estimation of variation in the X-Ray pixels are given in Eq. 5. We could thereby manipulate X-Ray for better 3D perception, using general CT or Atlas CT data while only using the X-Ray pixel values. In our experiments we use a clipping plane to interactively examine the X-Ray image (see Fig. 2).

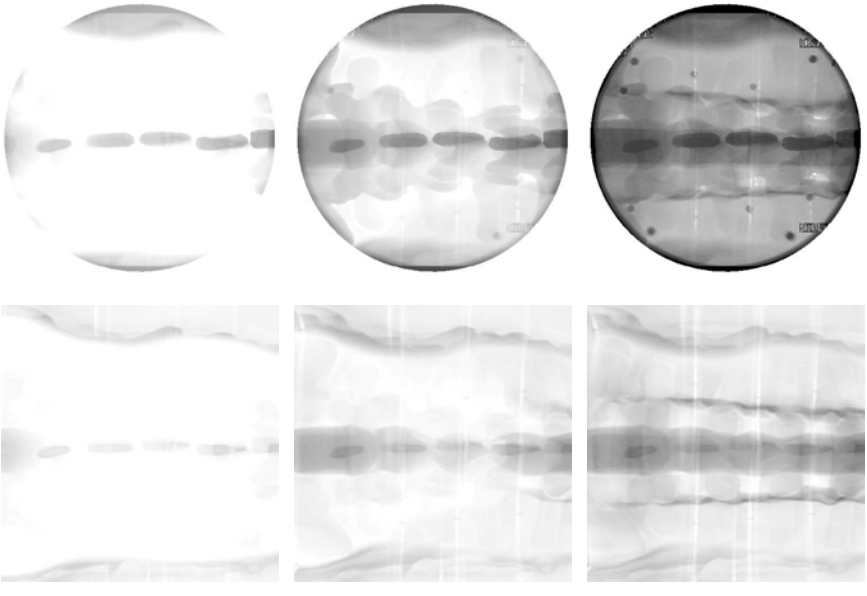


Fig. 2. Figures on top (round) show the result of IXPV of a clipping plane moving away from the viewer parallel to the image plane. Bottom figures (square) show corresponding clipped DRRs.

2.5 Fusion of Tools and X-Ray

Given a CAD model of a surgical tool and the possibility to track this tool, the virtual instrument can be directly fused with a given X-Ray image using IXPV and cutting the X-ray at the position where the virtual model of the tool is located. The results of our method are shown in Fig. 3 (left).

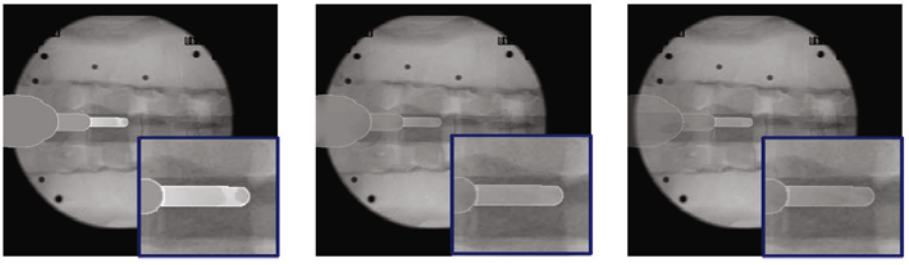


Fig. 3. A CAD model of a surgical drill fused with a single-view X-Ray image of a spine phantom. Compare IXPV (left) depth-of-field (center) and 2D blending (right).

2.6 Visualization of Co-registered X-Ray and Dynamic US

The visualization of fused X-Ray and ultrasound can take advantage of the depth cues given by IXPV. In this regard, the method implicitly takes occlusion related effects of the ultrasound plane and the X-Ray into account. For our experiments, a cow leg cadaver was positioned at the center of a transparent box with 6 reflective markers affixed to it that are visible in both CT and tracked NDI technology. Fusion between CT and ultrasound was realized using intensity based registration [6]. Having an accurate initial pose to an AP view of X-Ray, we fuse a B-mode ultrasound plane directly with X-Ray. Fig. 4 shows the effects of this fusion in comparison to 2D blending.

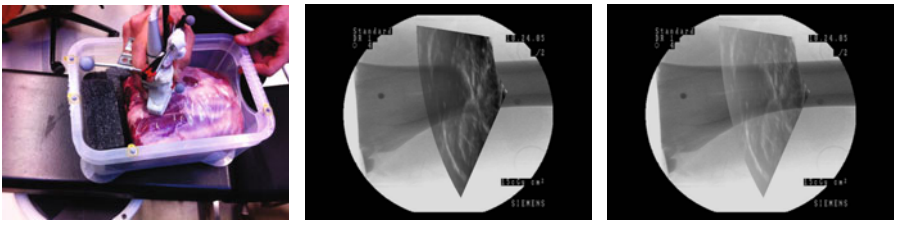


Fig. 4. Animal cadaver cow leg (left) visual assessment. (center) When applying the IXPV technique the tibia is displayed correctly in front of the ultrasound plane. Fusion using 2D blending of ultrasound and X-Ray (right) produces misleading depth cues.

2.7 Radiometric Error

Even if geometric 2D/3D registration of X-Ray and CT or Atlas could be done precisely thanks to recent advances in registration techniques, the estimation of the X-Ray intensity images is extremely difficult. This is due to 2D/3D registration optimizing a linear projection matrix while radiometric property estimation is a highly non linear procedure. The absorption value obtained from Eq. 5 is $\in [I_{\mathrm{XRay}}, \frac{I_{\mathrm{XRay}}}{I_{\mathrm{CT}}}]$. As for the difficulties of radiometric estimation: $\frac{I_{\mathrm{XRay}}}{I_{\mathrm{CT}}} \neq 1$.

To achieve reasonable visualization, (i.e. when integrating tracked instruments into X-ray using our method), we would like to have the object (i.e. instrument) fully visible in front of the anatomy. When moving behind anatomy only the X-Ray intensities should prevail. This intuition in mind we quantify the error in Eq. 5 as:

$$e^{-\int^{x_0} \mu(x)dx} - \frac{I_{\mathrm{XRay}}}{I_{\mathrm{CT}}} * e^{-\int^{x_0} \mu_{\mathrm{CT}}(x)dx} \tag{6}$$

Note that for $x_0 = 0$ (in front of the volume) this leads to: $1 - \frac{I_{\mathrm{XRay}}}{I_{\mathrm{CT}}}$. For $x_0 = 1$ this corresponds to no error. Thus we decided to add

$$\frac{x_{\max} - x}{x_{\max} - x_{\min}} * (1 - \frac{I_{\mathrm{XRay}}}{I_{\mathrm{CT}}}) \tag{7}$$

to Eq. 5 in order to get the described visual effects, where $x_{\max}$ and $x_{\min}$ would define the limits of our volume of interest along the ray.

2.8 Error from Registration in Depth

A common issue for 2D/3D registration algorithms is that they are least reliable in depth. Let δx be this error in depth. For IXPV the absorption function approximation (Eq. 5) becomes $\mu_{\mathrm{CT}}(x + \delta x) \approx \mu(x)$. Thus the error results in an offset within the intensity approximation and the error is given by:

$$e^{-\int^{x_0} \mu_{\mathrm{CT}}(x+\delta x)dx} - \frac{I_{\mathrm{XRay}}}{I_{\mathrm{CT}}} * e^{-\int^{x_0} \mu_{\mathrm{CT}}(x)dx} \approx e^{-\int^{x_0} \mu_{\mathrm{CT}}(x+\delta x)dx} - e^{-\int^{x_0} \mu_{\mathrm{CT}}(x)dx} \tag{8}$$

3 Results

The impact of our IXPV technique was validated through a series of phantom and cadaver experiments. For comparison, we also implemented 2D blending and depth-of-field techniques (see Fig. 3). The phantom study used co-registered CT and X-Ray of a spine phantom. A questionnaire was devised which is divided into two sections: depth perception in X-Ray and depth perception of clipping planes to evaluate the effect of interactivity. In total, eleven participants took part in the study (five experienced clinicians in the orthopedic and trauma surgery department and six senior researchers having a strong background in medical imaging). For section one (depth perception), we presented one top view composing a virtual model of a drill and the X-Ray, using either: (a) 2D blending (2D), (b) depth-of-field blending (dof) or (c) IXPV picture (compare Fig. 3). Participants were asked to choose correctly one out of three lateral views for the given top view. In a series of 27 questions, we randomized drill position depths (2, 5, or 10 millimeters), as well as the techniques.

3.1 User-Study

Participants received the following instructions for section two (X-Ray clipping): to choose correctly one of three lateral views given a static AP view and given videos of a clipping plane repeatedly moving through the object. In a series of 8 questions, we randomized between static image and video sequences. We accumulated statistical results of the questionnaire based on the total number of correct answers from all participants. Lastly, we asked only the surgeons to evaluate in a five-point Likert scale (i.e. 1 strongly disagree, 5 strongly agree), if IXPV was easier, more intuitive and more helpful for depth perception in X-Ray. A Wilcoxon rank sum test was used to calculate the p-values. *Depth perception:* The results for the three modes are presented in Fig. 5 (right). The p-value for IXPV vs. 2D blending turned out to be 0.0449 (significant). For IXPV vs. depth-of-field it is 0.0758 (not significant). *X-Ray interactive clipping:* The results are shown Fig. 5 (right). The p-value is 0.0119 (significant). *Likert scale:* The five surgeons responded positively to the IXPV method as suggested by a score of 3.7 ± 0.9.

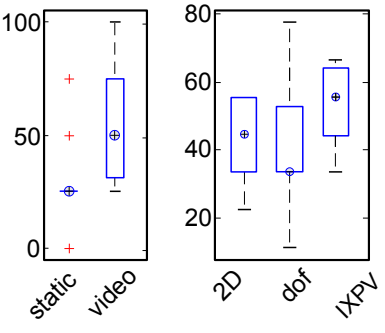


Fig. 5. Boxplots of the results

4 Discussion and Conclusion

This paper presents a novel technique of interactive X-Ray perceptual visualization (IXPV) which takes the physics of X-Ray into account to improve 3D perception within a single X-Ray image and allows fusion with tracked surgical tools or intraoperative ultrasound. Given an X-Ray image and a registered CT, our method relies on DVR algorithms to gain a modulation of original X-Ray intensities, to approximate those of a specific depth within the volume. Modern DVR algorithms and optimizations described in section 2.2 enable real-time visualization of IXPV even on low end graphics hardware. We evaluated our method using a survey which was designed to test two major hypotheses. Firstly, our method is superior to 2D blending. Although we were not able to reject the null hypothesis for IXPV vs. depth-of-fied blending, the results suggest, that IXPV provides more consistent depth cues than DOF. Secondly, our method enhances depth perception of 2D X-Ray images by allowing interactive examination with a clipping plane. Results suggest that IXPV should be further investigated and developed as a futuristic complimentary imaging tool to traditional radiographic images found in clinics. Future work includes the use anatomic models and/or of statistical atlas CT instead of patient specific data as well as fusion of IXPV with volumetric data, such as PET scans.

References

1. Boucheny, C., Bonneau, G.P., Droulez, J., Thibault, G., Ploix, S.: A perceptive evaluation of volume rendering techniques. In: Proceedings of the 4th Symposium on Applied Perception in Graphics and Visualization, APGV 2007, pp. 83–90. ACM, New York (2007)
2. Bruckner, S., Gröller, E.: Enhancing depth-perception with flexible volumetric halos. IEEE Transactions on Visualization and Computer Graphics 13, 1344–1351 (2007)
3. Chintalapani, G., Murphy, R., Armiger, R.S., Lepisto, J., Otake, Y., Sugano, N., Taylor, R.H., Armand, M.: Statistical atlas based extrapolation of ct data. In: Proceedings of SPIE, vol. 7625, p. 762539 (February 2010)
4. Díaz, J., Vazquez, P.: Depth-enhanced maximum intensity projection. In: IEEE/EG International Symposium on Volume Graphics, vol. 8, pp. 1–8 (2010)
5. Ebert, D., Rheingans, P.: Volume illustration: non-photorealistic rendering of volume models. In: Proceedings of the Conference on Visualization, VIS 2000, pp. 195–202. IEEE Computer Society Press, Los Alamitos (2000)
6. Fallavollita, P., Aghaloo, Z.K., Burdette, E.C., Song, D.Y., Abolmaesumi, P., Fichtinger, G.: Registration between ultrasound and fluoroscopy or ct in prostate brachytherapy. Medical Physics 37(6), 2749–2760 (2010)
7. Gong, R.H., Stewart, J., Abolmaesumi, P.: Reduction of multi-fragment fractures of the distal radius using atlas-based 2d/3d registration. In: Proceedings of SPIE. Medical Imaging 2010: Visualization, Image-Guided Procedures, and Modeling, Orlando, Florida, USA, vol. 7261, p. 726137 (February 2009)
8. Kersten, M., Stewart, J., Troje, N., Ellis, R.: Enhancing depth perception in translucent volumes. IEEE Transactions on Visualization and Computer Graphics 12, 1117–1124 (2006)
9. Khamene, A., Bloch, P., Wein, W., Svatos, M., Sauer, F.: Automatic registration of portal images and volumetric CT for patient positioning in radiation therapy. Medical Image Analysis 10(1), 96–112 (2006)
10. Markelj, P., Tomazevic, D., Likar, B., Pernus, F.: A review of 3d/2d registration methods for image-guided interventions. Medical Image Analysis (2010)
11. Svakhine, N.A., Ebert, D.S.: Interactive volume illustration and feature halos. In: Proceedings of the 11th Pacific Conference on Computer Graphics and Applications, PG 2003, p. 347. IEEE Computer Society, Washington, DC (2003)
12. Svakhine, N.A., Ebert, D.S., Andrews, W.M.: Illustration-inspired depth enhanced volumetric medical visualization. IEEE Transactions on Visualization and Computer Graphics 15, 77–86 (2009)

Augmented Reality System for Oral Surgery Using 3D Auto Stereoscopic Visualization

Huy Hoang Tran[1], Hideyuki Suenaga[2], Kenta Kuwana[1], Ken Masamune[1], Takeyoshi Dohi[1], Susumu Nakajima[2], and Hongen Liao[1]

[1] The University of Tokyo, Tokyo 113-8656, Japan
[2] The University of Tokyo Hospital, Tokyo 113-8655, Japan

Abstract. We present an augmented reality system for oral and maxillofacial surgery in this paper. Instead of being displayed on a separated screen, three-dimensional (3D) virtual presentations of osseous structures and soft tissues are projected onto the patient's body, providing surgeons with exact knowledge of depth information of high risk tissues inside the bone. We employ a 3D integral imaging technique which produce motion parallax in both horizontal and vertical direction over a wide viewing area in this study. In addition, surgeons are able to check the progress of the operation in real-time through an intuitive 3D based interface which is content-rich, hardware accelerated. These features prevent surgeons from penetrating into high risk areas and thus help improve the quality of the operation. Operational tasks such as hole drilling, screw fixation were performed using our system and showed an overall positional error of less than $1mm$. Feasibility of our system was also verified with a human volunteer experiment.

1 Introduction

Image guided systems have been supporting surgeons in a wide range of clinical applications from diagnostic to postoperative evaluation [1,2]. In oral and maxillofacial surgery, the prerequisite for a successful operation is the exact knowledge of high risk tissues such as blood vessels and dental nerves so that these inferior tissues can be avoided [3]. One challenging problem in current navigation systems is the lack of depth information of the visualized structures due to the limitation of conventional flat displays. In addition, as the visual information is often separated from the surgical site, surgeons have to make a hand-eye transformation in order to match visual information to the real scene.

In order to give the surgeons an intuitive grasp of the depth information of a surgical scene, several approaches have been presented [4,5]. Among those, the image overlay system using Integral Videography (IV) [6,7] is a promising solution for the simplicity in implementation and the ability to produce motion parallax in both horizontal and vertical directions. In such system, users can see the 3D internal structure superimposed onto the actual anatomy through a half-silvered mirror [8]. However, the low rendering speed and the lack of the ability to visualize enriched information make it difficult to use such system in navigation.

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 81–88, 2011.

For example, one would expect to interactively change the transparency of a surface to see the blood vessel inside, or get a real-time visual feedback about the progress of the operation.

In this study, we present an augmented reality system for oral surgery using IV image. We develop a GPU based rendering algorithm to accelerate the IV images generation. Our system also offers an intuitive 3D based user interface that allows surgeons to interact with the image in 3D environment. Feasibility study of the system is carried out with a set of experiments including a human volunteer experiment.

2 Materials and Methods

2.1 3D IV Display and Rendering

The basis of IV comes from the principle of Integral Photography [9]. To reproduce 3D image of an object in real space, a micro lens array is placed in front of a high density LCD display. Each pixel in the portion behind a micro lens emits a light ray in a unique direction that connects the pixel and the center of the lens. From a specific viewpoint, one can only see the set of light rays emitted from the pixels that lie on the straight lines connecting the viewpoint and the lenses. Consequently, different aspects of an object can be observed from different directions, giving the observer a sense of depth(Fig. 1(a)).

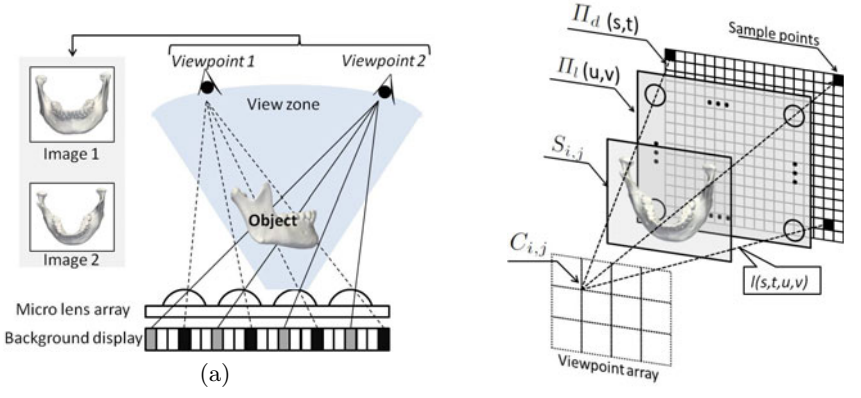


Fig. 1. (a) Principle of Integral Videography. (b) Sampling an image using Π_l and Π_d

In order to create an IV image, every pixel on the background display need to be computed. There are two main IV rendering methods: volume ray-tracing and surface based rendering. In the former, value of each pixel is depends on how the corresponding light ray intersects with the object. This method requires computation for every light ray but creates very accurate 3D images.The latter uses CG surface rendering to render an array of multi-view images from a set

of predefined viewpoints. These images are then mapped backward to the background image using pixel distribution algorithm [10]. This method has all the advantages of surface rendering (shading, texture mapping, etc.). However, due to the fact that the spatial arrangement and shape of the lens in general differ from those of the pixels in an image, apparent distortions are observable. For that reason, surface based method is often used for other purposes than surgical navigation.

In order to render accurate IV image while maintaining the desired visual effects, we introduce a light field [11] approach to the surface based method. The image $S_{i,j}$ rendered from view $C_{i,j}$ is projected onto the lens array plane Π_l and the display plane Π_d with the center of projection located at $C_{i,j}$. Applying the transformation for all i and j will result in a densely sampled light space in which each light ray is defined as a four parameters function $l(s,t,u,v)$ ((s,t) from Π_d and (u,v) from Π_l)(Fig. 1(b)).

Value of a pixel located at $\mathbf{P} = (s,t)^T$ in Π_d is computed as follow.

$$I(s,t) = \int\int_{\Pi_l} \Theta(s,t,u,v) l(s,t,u,v)\, \mathrm{d}u\mathrm{d}v, \tag{1}$$

where $\Theta(s,t,u,v)$ is a weight function that defines the contribution of the light ray to the final result such that

$$\int\int_{\Pi_l} \Theta(s,t,u,v)\, \mathrm{d}u\mathrm{d}v = 1. \tag{2}$$

The design of the weight function is beyond the scope of this paper. However, for a well-manufactured lens array, contribution of light rays from lenses located far away from the query pixel can be neglected. Therefore a Dirac delta function in which the base width equals to the diameter of one lens can be a good choice.

$$\Theta(s,t,u,v) \stackrel{def}{=} \delta_R(d) = \frac{1}{R\sqrt{\pi}} e^{-\frac{d^2}{R^2}}, \tag{3}$$

where d is the Euclidean distance between $(s,t)^T$ and $(u,v)^T$, R is the radius of a micro lens.

$l(s,t,u,v)$ can be computed by resampling Π_l and Π_d, this is done through several interpolations using the sampled points. These interpolations are in general time consuming tasks that slow down the rendering speed significantly. Fortunately, recent GPUs provide a very effective way to perform these interpolations through their naive texture accessing function. To do so, all of the images are rendered and stored in the GPU's texture memory. Next, a rectangle is rendered as a proxy geometry where textures are loaded. At the final stage of the graphics pipeline, each fragment (i.e. a potential pixel) is fed with the pixel value from equation 1. Per fragment processing is implemented using a GPU shading language, the C for Graphics [12]. Our algorithm does not require external CPU memory so we can get rid of the bottlenecks in data transmission between GPU and CPU.

2.2 IV Augmented Reality System

A half-silvered mirror is attached to the IV display. Through the half mirror, surgeons can see the reflected IV image superimposed onto the patient (Fig. 2(a)). We use a fly eyes type lens array with lens diameter of 1.001 mm, which is fixed to a 6.4 inches high resolution LCD (203 dpi). In order to track the movement of the patient, a marker frame is fixed to the lower jaw. The marker frame is calibrated using a model of exactly the same size with the lower jaw.

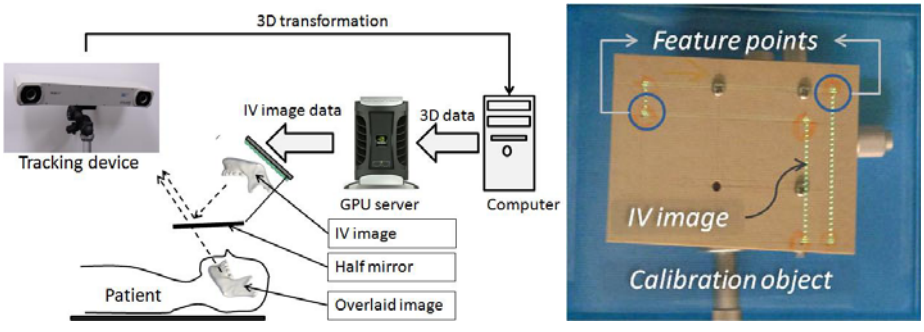


Fig. 2. (a) IV augmented reality system configuration. (b) IV image overlaid with calibration object.

Although the geometry of the IV display-half mirror combination is known, a small displacement of the lens array with respect to the display would result in a large displacement of the IV Image. To accurately register the IV image to the world coordinate system, we use a calibration object which consists of six feature points. A model of the object is created inside a computer and its IV image is then overlaid with the real object. Manual adjustment in position and orientation is made to the object until all of the feature points are aligned completely as shown in Fig. 2(b).

In order to support surgeons in operational tasks such as hole drilling, bone resection and screw fixation, in addition to the osseous structures, IV images of the instrument, surgical path, soft tissues are utilized in our system. Surgeons can verify the progress of the operation, i.e. current position of the instrument, status of the bone and the risk tissues visually or through text based information. The reason why we add text based information is that it improves the surgeons' confidence in situations where small differences (less than 1mm) are difficult to recognize visually.

3 Experiments and Results

3.1 Image Generation Speed

The prerequisite for our system is that movement of the instrument as well as the patient has to be updated in real-time. In addition, surgeons should have the

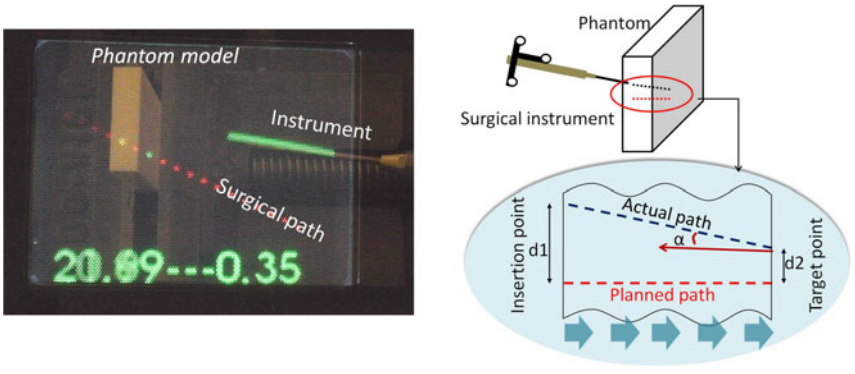


Fig. 3. Drill experiment using image overlay system

ability to change the properties of the displayed objects to obtain the desired visual effect. Therefore, we evaluated the rendering speed of each individual object. The specifications of our computer system are as follow.

CPU: Pentium Core 2 Duo, 2.4 GHz
GPU: NVIDIA Quadro Plex model 4 (Quadro FX 5600×2)
Elemental IV image size: 1024x768 pixels

The rendering system can render IV image of a surgical instrument at 80 frames per second (fps). For a complex anatomical model, e.g. the whole lower jaw, rendering speed is 5 fps. Therefore, the minimum updating speed of our system is around 5 fps. In most cases, our system shows a rendering speed of more than 10 fps, which is sufficient for surgical navigation.

3.2 Accuracy Evaluation

In order to evaluate the accuracy of our system when used in a real clinical case, a surgeon was asked to performed operational tasks such as hole drilling using real surgical instruments. The drilling task was performed on a phantom model imitating a bone surface. For the sake of simplicity in error estimation, simple models with known geometry were used rather than complex anatomical models. The objective of the operation is to drill through the model along a predefined path. During the operation, the surgeon was asked to keep the instrument's axis as close as possible to the surgical path, otherwise the operation would be considered a failure.

The drill path was set to be not perpendicular to the model's surface so that the surgeon would not get a hint from the model's surface. As shown in (Fig. 3), the drill path was inclined at approximately 45° to the model's surface.

When the instrument is aligned properly to the surgical path, color of the instrument is automatically changed to alert the surgeon that it is safe to proceed. The surgeon will then start his drilling task.The operation ends when the drill

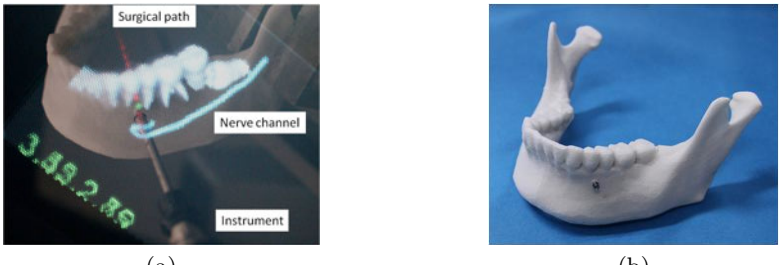
reaches the target point on the other side of the model. By measuring the position of the drill holes in both sides, we can compute the positional and angular errors of the operation.

After 10 trials, positional error of the insertion point (d_1) was $0.7 \pm 0.4mm$, the error of the target point (d_2) was $0.5 \pm 0.3mm$ and the angular error (α) was $3.3 \pm 1.1°$. Error of the target tends to be smaller than that of the insertion point. The reason is that even the first approach to the model's surface was not good, the surgeon was still able to correct the orientation of the instrument during the drilling task.

3.3 Feasibility Evaluation

Screw fixation using model. Another task that is often performed during oral surgery is screw fixation. We use the anatomical model of a lower jaw in this experiment. The prerequisite for a successful operation is that blood vessels and nerves must not be damaged. For that reason, a nerve channel underneath the bone surface is highlighted (Fig. 4(a)). During operation, an alert is sent to the surgeon when the instrument matches the planned path, or when the nerve channel is likely to be touched. Again, a surgeon was asked to perform the screw fixation under the guidance of our system. In order to mimic as closely as possible the clinical situation, the surgeon used the instrument set (screw and screw driver) that was manufactured for medical use.

After the fixation was completed, we verified that the screw was fixed in the proper place as planned in the planning stage, with no threat of damaging the nerve (Fig. 4(b)).



(a) (b)

Fig. 4. (a) IV image of the bone with blood vessel highlighted. (b) screw fixation result.

Human volunteer experiment. A problem should come up when utilizing our system in real clinical case is the limited work space due to the presence of IV display and half mirror in the surgical area. Another problem should be the response speed of the system when patient make a movement. A volunteer experiment was carried out by the same surgeon to evaluate the influence of those factors in real clinical applications.

Fig. 5 (left) shows the settings of this experiment. The half mirror was placed at about 25 cm above the surgical area, giving the surgeon a sufficient work space. A marker frame was attached tightly to the patient's lower jaw for a proper position tracking. Fig. 5 (right) shows the surgical area as viewed from the surgeon's point of view. The IV image was superimposed with the patient correctly, giving the surgeon a fast and accurate understanding of the surgical area. Any movement made by the patient (movement of the head, the lower jaw, etc.) was properly detected and IV images are updated with a short time delay.

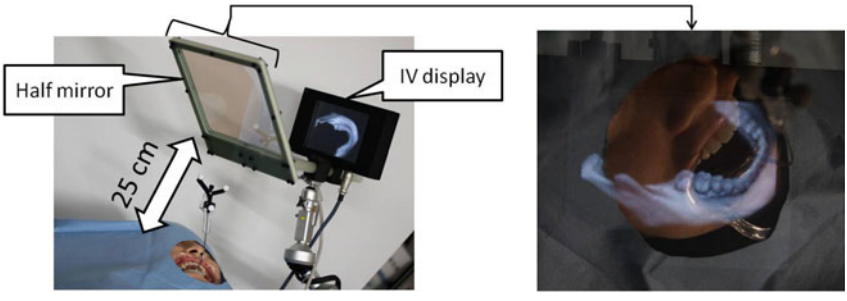


Fig. 5. Settings of volunteer experiment (left), IV image overlaid onto surgical area (right)

4 Discussion and Conclusion

In our experiments, IV image gave surgeons a fast and intuitive estimation of the target structures in real space. However, as each micro lens becomes a logical pixel, when the instrument was closed to the surgical path (the distance is smaller than one lens), surgeons tended to rely more on textual information than visual information because displacement smaller than one lens was difficult to recognize. For that reason, the system proposed in this paper has better accuracy than previously developed systems ($0.7 \pm 0.4mm$). This accuracy met the requirement for use in clinical application and thus the potential of applying our system in real clinical cases has been proven. In addition, the GPU based rendering algorithm has enabled real-time tracking of the patient movement. Although a short time delay can be noticed, this would not become a problem because the patient movement is not so dynamic during operation.

Despite the installation of the IV display and the half mirror, surgeons still have sufficient work space for the surgery. Moreover, it can be easily removed from the surgical area when AR representation is not needed. This feature makes our system very flexible and pose no restriction to the operation workspace.

In conclusion, we have developed an augmented reality system to support surgeons in oral and maxillofacial surgery. Our system provides an on-site,fast and accurate 3D virtual representation of the surgical area. A number of experiments mimicking real clinical interventions and showed promising results. In the

future, by improving some of the technical aspects, particularly the registration method, we intend to carry out clinical trials on real patients.

Acknowledgment. This work was supported in part by Research Fellowship of the JSPS (21.9746), Grant for Industrial Technology Research (07C46050) of NEDO, Japan, Grant-in-Aid for Scientific Research on Innovative Areas (22103505) of the MEXT in Japan, and KAKENHI (222400061, 22659366).

References

1. Germano, I.M., Villalobos, H., Silvers, A., Post, K.D.: Clinical use of the optical digitizer for intracranial neuronavigation 45, 261–269 (1999)
2. Plaweski, S., Cazal, J., Rosell, P., Merloz, P.: Anterior cruciate ligament reconstruction using navigation: a comparative study on 60 patients. The American Journal of Sports Medicine 34, 542–552 (2006)
3. Hassfeld, S., Muehling, J.: Computer assisted oral and maxillofacial surgery: A review and an assessment of technology. Int. Oral Maxillofac Surg. 30, 2–13 (2001)
4. Blackwell, M., Nikou, C., M.DiGioia, A., Kanade, T.: An Image Overlay system for medical data visualization. Medical Image Analysis 4, 67–72 (2000)
5. Traub, J., Sielhorst, T., Heining, S.M., Navab, N.: Advanced Display and Visualization Concepts for Image Guided Surgery. Journal of Display Technology 4, 483–490 (2008)
6. Liao, H., Nakajima, S., Iwahara, M., Kobayashi, E., Sakuma, I., Yahagi, N., Dohi, T.: Development of Real-Time 3D Navigation System for Intra-operative Information by Integral Videography. Journal of Japan Society of Computer Aided Surgery 2, 245–252 (2000)
7. Liao, H., Iwahara, M., Hata, N., Dohi, T.: High-quality integral videography using a multi-projector. Optics Express 12, 1067–1076 (2004)
8. Liao, H., Inomata, K., Sakuma, I., Dohi, T.: 3-D Augmented Reality for MRI-guided Surgery using Integral Videography Autostereoscopic-image Overlay. IEEE Transactions on Biomedical Engineering 57, 1476–1486 (2010)
9. Lippmann, M.G.: Epreuves reversibles donnant la sensation du relief. J. de Phys. 7, 821–825 (1908)
10. Liao, H., Nomura, K., Dohi, T.: Autostereoscopic Integral Photography Imaging using Pixel Distribution of Computer Graphics Generated Image. In: ACM SIGGRAPH (2005)
11. Levoy, M.: Hanrahan, p.: Light field rendering. In: Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques, pp. 31–42 (1996)
12. Fernando, R., Kilgard, M.J.: The Cg Tutorial. Addison-Wesley Professional, Reading (2003)

Dense Surface Reconstruction for Enhanced Navigation in MIS

Johannes Totz, Peter Mountney, Danail Stoyanov, and Guang-Zhong Yang

The Hamlyn Centre for Robotic Surgery
Imperial College London, UK

Abstract. Recent introduction of dynamic view expansion has led to the development of computer vision methods for minimally invasive surgery to artificially expand the intra-operative field-of-view of the laparoscope. This provides improved awareness of the surrounding anatomical structures and minimises the effect of disorientation during surgical navigation. It permits the augmentation of live laparoscope images with information from previously captured views. Current approaches, however, can only represent the tissue geometry as planar surfaces or sparse 3D models, thus introducing noticeable visual artefacts in the final rendering results. This paper proposes a high-fidelity tissue geometry mapping by combining a sparse SLAM map with semi-dense surface reconstruction. The method is validated on phantom data with known ground truth, as well as in-vivo data captured during a robotic assisted MIS procedure. The derived results have shown that the method is able to effectively increase the coverage of the expanded surgical view without compromising mapping accuracy.

1 Introduction

In Minimally Invasive Surgery (MIS), disorientation due to restricted field-of-view of the endoscopic camera has a recognised effect on 3D visuomotor control and navigation. Dynamic View Expansion (DVE) combines previously captured images with the current view of the camera, thus revealing the surrounding anatomical structure to minimise these effects. In practise, DVE can be performed using image mosaicing algorithms. Such algorithms register two or more images together based on the assumption of the 3D structure of the scene. The most common assumption is that the tissue or organ is planar and the registration is performed by using image- or feature-based techniques. The former generally requires a large overlap between the images, whereas feature-based techniques [1] are based on the detection and matching of salient regions in the images. In reality, tissue structure and organ geometry in MIS tend to not satisfy the planar assumptions. To alleviate this effect, parallax motion correction has been proposed [2] and prior knowledge of the organ geometry has been used to model intraluminal structures such as the colon [3] and oesophagus [4] as cylinders. However, if the assumed tissue model does not represent the tissue geometry, significant artefacts can be introduced into the final result with expanded field-of-view. In recent work based on Simultaneous Localisation and

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 89–96, 2011.

Mapping (SLAM), methods are proposed to build a sparse model of the tissue structure sequentially and incrementally without relying on prior knowledge or explicit assumption of the tissue structure [5,6]. The sparse model is meshed and textured with images from the endoscope. This enables the model to be projected onto the estimated camera position to create an expanded view. With this approach, two problems have been identified due to the sparse representation of the tissue. Firstly, artefacts are introduced into the visualisation due to the coarse model. Secondly, the expanded view is limited to the size of the modelled region. Any part of the tissue that does not fall within the model cannot be visualised with DVE. In [7], an approach for dense surface reconstruction is used to create 3D tissue models from stereo laparoscopic images. The method matches a sparse set of salient regions using stereo Lucas-Kanade and propagates the disparity around each matched region. The method is used to reconstruct deforming tissue from a static camera. Dense surface reconstruction can also be achieved using monocular images combined with [8] or without [9] the use of a robotic arm. However, limited work has been performed in dealing with a dense model of the tissue from a moving stereo laparoscopic camera.

The purpose of this paper is to explore the use of sparse SLAM with dense surface reconstruction for DVE. Under the proposed framework, the disc-homeomorphic 3D model is parameterised into a planar texture domain into which video images and surface details are projected. The accuracy of the proposed method is validated on phantom data and the potential clinical application is demonstrated on in-vivo data.

2 Methods

Fig. 1 outlines the main steps involved in creating the expanded visualisation. SLAM is used to estimate the position of the camera and incrementally build a sparse map or model of the tissue. This model is extended using a "skydome" and a dense stereo algorithm is used to reconstruct the surface of the tissue, which is registered to the sparse model. The dense tissue model is then textured and reprojected onto the live laparoscopic video stream to create an augmented visualisation. It should be noted that the original live video stream is never modified to ensure the visual fidelity of the current surgical view.

2.1 Dynamic View Expansion with Sparse Stereo SLAM

For sequential scene reconstruction, SLAM is used to estimate the position and orientation of the stereo camera and a sparse set of features in 3D which represent the map or tissue model. This information comprises the state of the system which is modelled probabilistically using an Extended Kalman Filter (EKF), enabling uncertainty and noise in the system to be modelled. SLAM detects and matches regions of interest in the stereo images, which are triangulated to estimate a 3D map feature. In subsequent video frames, the motion of the camera is predicted using a constant velocity, constant acceleration motion model. The

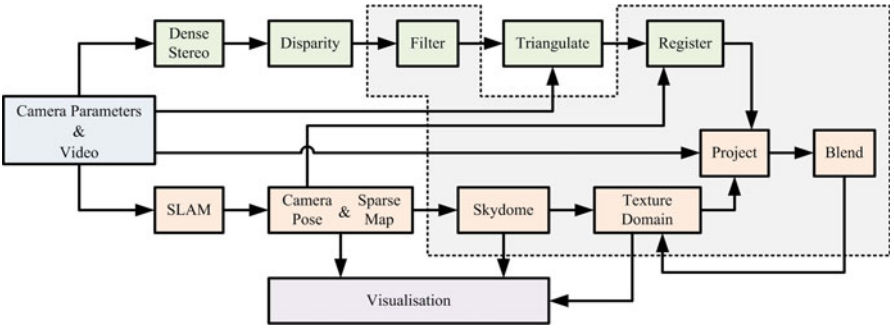


Fig. 1. Processing pipeline of the proposed method. Video images are processed independently by SLAM and dense-disparity until they are registered to each other. Projection blends surface detail and video colour into texture maps used for visualisation. This paper addresses mainly components enclosed in the dashed box.

position of each 3D map feature is estimated relative to the predicted camera position and the map features are measured by matching the regions of interest in the image space. The prediction and measurement are used to perform the update step of the EKF and estimate a new state (camera position and 3D map). A photorealistic model of the tissue is generated by meshing the sparse 3D map using a Delaunay triangulation and texturing using the approach outlined in [6].

Using the SLAM-map alone for DVE can result in a loss of visual information because parts of the video frame may not be covered with trackable features and map triangles. Previously, this was avoided by introducing virtual features into the map. For the proposed method, the SLAM-map triangulation Y is padded with additional triangles around the boundary to produce Y', which is referred to as "skydome" in this paper. The 3D position of vertices is computed by $y'_i = r \cdot (y_i - c) + c$, with c being the centroid of all n map vertices, y_i being a vertex on the boundary of the map and y'_i the position of the new skydome vertex. The factor r controls by how much the triangles are enlarged; suitable values are between 2 and 4, trading coverage for texture map resolution. Fig. 2a illustrates this aspect of the algorithm. The position of y' in the texture domain is derived in the same way, using 2D texture coordinates instead of 3D positions. In Figs. 2b and c, the green area marks video content covered by the sparse map only, while the red area marks additional video pixels covered by the skydome. Thus, the skydome provides a larger target for video frames to be mapped on; in fact, its exact geometry is not important because dense 3D positions mapped on it (Sec. 2.2) override its sparse vertex positions (Fig. 3d and e).

2.2 Dense Surface Reconstruction

Using the stereo images of the observed tissue, the 3D position of each pixel is computed by first establishing a disparity map that sets up correspondences between pixels in the left and the right video frame. The algorithm used to

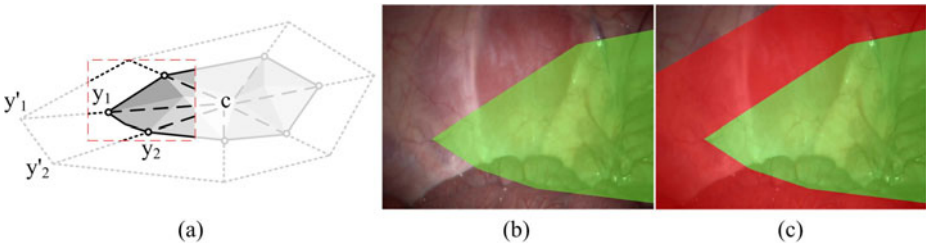


Fig. 2. Padding the SLAM-map with additional triangles. b) and c) illustrate the effect in the video image with green being the original sparse SLAM-map and red the additional area covered by the skydome.

compute the disparity between the left and right frames is the one proposed in [7]. Given a set of sparse correspondences between left and right images used as seeds, the algorithm starts by propagating for any neighbouring pixels at which a match in the other image can be found. The output is the disparity map D (Fig. 3b): for every pixel in the left image it contains a coordinate into the right image for the corresponding pixel, or an invalid coordinate if no such correspondences can be established. The outliers and regions without stereo-disparity correspondence are then filled by using Telea's in-painting method [10] followed by a box-blur. Figs. 3a-c show an example of the left and right video frame, the unprocessed disparity with holes and outliers, followed by in-painting. Each pixel in the disparity map is triangulated into 3D space. The reconstructed surface is mapped from the camera space to the SLAM coordinate system by transforming it with the SLAM-estimated camera pose. The remaining difference is due to SLAM's noise modelling, but mainly limited to the viewing direction. Therefore, for every SLAM feature visible in the current frame, the offset between its current position in 3D and the dense estimated surface is computed, subject to outlier removal. The resulting 3D position map P in the SLAM-coordinate system is rendered into the texture domain T already used for the video pixels [6], i.e. every video pixel that has been observed previously has a position in 3D.

2.3 Texture Domain Update and Visualisation

Within the proposed framework, updating texture domain T consists of three steps: reprojection, position update and colour update. There are three images attached to it: colour from video T_C, age-map T_A (storing a pixel's age, see [6]) and 3D position T_{3D} encoded in RGBA. The contents of these images is first reprojected from the previous frame to accommodate the possibly updated and retriangulated SLAM-map. Colour and position are normalised by the accumulated value in the alpha channel: $g_{rgba} = g_{rgba}/g_a$, where g is a pixel's RGBA colour value representing either colour or position.

Following the registration described above, the 3D position image is updated by rendering the sparse map Y' textured with P into T_{3D} with additive blending,

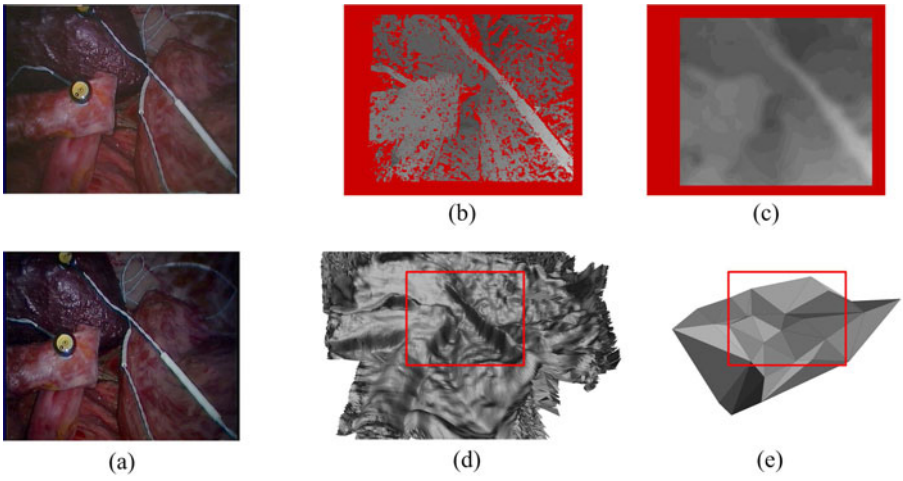


Fig. 3. Stereo disparity and estimated dense surface. a) shows left and right video frame of the phantom sequence; b) shows raw stereo disparity; c) filtered and smoothed disparity; d) shows the flat-shaded highly tesselated Y' with the red rectangle marking the part that would be covered by the video frame. For comparison, the sparse SLAM map surface Y is shown in e).

combining new and existing data (in effect dense temporal surface reconstruction). The colour and age-map are updated by taking the up-to-date detailed surface position from T_{3D} into account and each pixel is warped to account for perspective distortion. This two-pass approach is necessary to avoid an unstable feedback loop between the current and the previous frame's positions.Without normalisation during texture domain reprojection, newly arriving data would have minimal effect as many more frames have already been accumulated. Instead, a running average is computed. In addition, video and position projection is radially attenuated to account for spot-light illumination effects by applying a linear fall-off function towards the edges of the video frame.

The final result for each frame is visualised by using the fade-to-grey method of [6] but additionally rendering Y' with high tessellation and displacement mapping. The colour and position information are texture-mapped onto the triangles and then subdivided by at least eight levels using the graphics hardware. Each vertex is positioned in 3D space using the estimated dense surface position. The surface is never explicitly meshed; Fig. 3d serves for illustration purposes only. Pixels are rejected for which attenuation has caused $g_a < 1$ accumulation to avoid noisy edge cases.

3 Results

The method has been evaluated on a phantom sequence and an in-vivo porcine experiment, both representing an exploration of the abdomen. The stereo cameras were calibrated at the start and remain unchanged during the experiments.

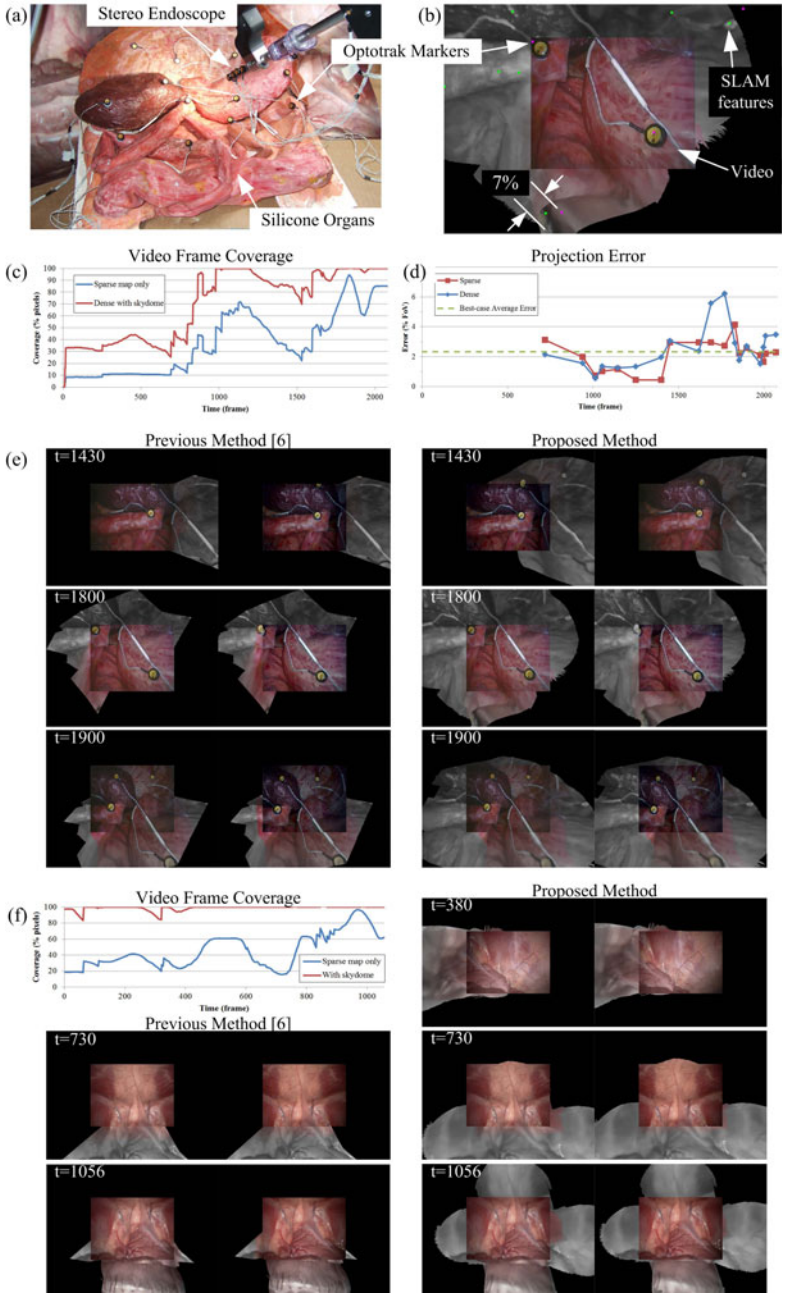


Fig. 4. a) Experimental setup with silicone organs, OptoTrak markers and stereo laparoscope. b) Projected markers (purple) and SLAM features (green) in the DVE. c) Video frame coverage of SLAM-map with and without skydome. d) Projection error of markers into image plane. e) Stereoscopic DVEs for phantom sequence at various points in time, comparing sparse with dense surface. f) Coverage plot and DVEs for in-vivo sequence.

The abdominal phantom used for validation comprised a full set of silicone organs painted with acrylic paint to provide realistic texture. Optical markers were placed on the surface of the phantom (see Fig. 4a) to provide the ground truth data represented as a sparse 3D model of the phantom. A further eight markers were placed on a rigid body attached to the laparoscope. These were used, in conjunction with hand-eye calibration, to determine the position of the camera. The quality of DVE was validated by projecting the position of the markers on the model into the expanded image plane of the camera at a given camera position. This provided a set of ground truth points in the 2D image plane for comparison with the reprojected dense model obtained using the proposed method. Due to the limited accuracy of the optical tracker and noise in the calibration the validation framework was evaluated to be accurate to within 2.3% of the FoV. The graph in Fig. 4d shows the average reprojection error for the sparse (red) and dense (blue) approaches. The average visual angel error for the sparse is 2.2% and for the dense 2.7% of the FoV, making the accuracy of the two results comparable. The accuracy of the dense method decreases around frame 1700 due to noisy estimation of the camera position causing inaccurate registration of the dense map but recovers afterwards. Qualitative evaluation of the sparse and dense approaches are provided in Fig. 4e. The effect of ageing the expanded view is clearly visible on these sequences with older regions appearing desaturated.

Quantitative evaluation is also performed with respect to coverage measured by how much of the observed tissue is represented in the expanded view. It is computed as the number of video frame pixels covered by a projection of the sparse map Y'. This correlates directly with how much information in the video is actually used for DVE: only video pixels that map to a position in the sparse map (with or without skydome) can be used. Fig. 4c depicts the effect of extending sparse map coverage via the skydome with $r = 2$. The phantom sequence starts with small rotational motion that only covers a relatively small area and does not introduce new SLAM features. Only after frame 700 does the camera move enough for new features to be introduced. The average video pixel coverage without skydome is 37% and with skydome 70%. The in-vivo sequence [11] shows the peritoneal wall, large intestines and stomach. The sequence is quantitatively evaluated in Fig. 4e and the average coverage without and with skydome is 45% and 99%, respectively, for $r = 3.5$ (chosen due to initial SLAM feature distribution). Easily visible in the coverage graph are the introduction of new SLAM features and the resulting extension of the skydome: sudden spikes in increased coverage. Fig. 4e shows the video frames with the corresponding DVE. The current single-threaded implementation of the proposed method runs at about 1 frame per second.

4 Conclusions

In this paper, a method has been proposed for DVE by exploiting the spatial-temporal correspondence of intra-operative laparoscopic imaging. The novelty of the proposed approach is in the creation of a dense tissue model from a moving stereoscopic camera and the method is validated on both silicon phantom

with know ground truth and in-vivo data. The proposed method has shown to increase the coverage of the expanded view without compromising accuracy. Another feature of the proposed system is that the method is sequential and well suited for real-time operation. In addition the system has been designed to make it compatible with non-static SLAM algorithms and dense non-rigid tissue models. It is worth noting that noise modelling in SLAM takes into account unreliable feature tracking thus increasing the robustness of the method. Further improvements of the proposed method could include estimation of the surface's BRDF for perspective correct relighting.

References

1. Hu, M., Penney, G., Rueckert, D., Edwards, P., Bello, F., Figl, M., Casula, R., Cen, Y., Liu, J., Miao, Z., Hawkes, Z.M.D.: A robust mosaicing method with super-resolution for optical medical images. In: Liao, H., Edwards, P.J., Pan, X., Fan, Y., Yang, G.-Z. (eds.) MIAR 2010. LNCS, vol. 6326, pp. 373–382. Springer, Heidelberg (2010)
2. Lerotic, M., Chung, A.J., Clark, J., Valibeik, S., Yang, G.-Z.: Dynamic view expansion for enhanced navigation in natural orifice transluminal endoscopic surgery. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 467–475. Springer, Heidelberg (2008)
3. Seshamani, S., Lau, W., Hager, G.D.: Real-time endoscopic mosaicking. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4190, pp. 355–363. Springer, Heidelberg (2006)
4. Carroll, R.E., Seitz, S.M.: Rectified surface mosaics. In: IEEE 11th International Conference on Computer Vision, pp. 1–8 (October 2007)
5. Mountney, P., Yang, G.Z.: Dynamic view expansion for minimally invasive surgery using simultaneous localization and mapping. In: 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 1184–1187 (September 2009)
6. Totz, J., Fujii, K., Mountney, P., Yang, G.Z.: Enhanced visualisation for minimally invasive surgery. International Journal of Computer Assisted Radiology and Surgery (2011), doi:10.1007/s11548-011-0631-z
7. Stoyanov, D., Scarzanella, M.V., Pratt, P., Yang, G.-Z.: Real-time stereo reconstruction in robotically assisted minimally invasive surgery. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6361, pp. 275–282. Springer, Heidelberg (2010)
8. Moll, M., Tang, H.-W., Van Gool, L.: GPU-accelerated robotic intra-operative laparoscopic 3D reconstruction. In: Navab, N., Jannin, P. (eds.) IPCAI 2010. LNCS, vol. 6135, pp. 91–101. Springer, Heidelberg (2010)
9. Newcombe, R.A., Davison, A.J.: Live dense reconstruction with a single moving camera. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1498–1505 (2010)
10. Telea, A.: An image inpainting technique based on the fast marching method. Journal of Graphics Tools 9(1) (2004)
11. Mountney, P., Stoyanov, D., Yang, G.Z.: Three-dimensional tissue deformation recovery and tracking: Introducing techniques based on laparoscopic or endoscopic images. IEEE Signal Processing Magazine 27(4), 14–24 (2010)

Compressed Sensing Based 3D Tomographic Reconstruction for Rotational Angiography

Hélène Langet[1,2,3], Cyril Riddell[1], Yves Trousset[1], Arthur Tenenhaus[2], Elisabeth Lahalle[2], Gilles Fleury[2], and Nikos Paragios[3]

[1] GE Healthcare, 78530 Buc, France
[2] Supélec, Department of Signal Processing and Electronic Systems, 91192 Gif-sur-Yvette, France
[3] Ecole Centrale Paris, Applied Mathematics and Systems Department, 92290 Châtenay-Malabry, France

Abstract. In this paper, we address three-dimensional tomographic reconstruction of rotational angiography acquisitions. In clinical routine, angular subsampling commonly occurs, due to the technical limitations of C-arm systems or possible improper injection. Standard methods such as filtered backprojection yield a reconstruction that is deteriorated by sampling artifacts, which potentially hampers medical interpretation.

Recent developments of compressed sensing have demonstrated that it is possible to significantly improve reconstruction of subsampled datasets by generating sparse approximations through ℓ_1-penalized minimization. Based on these results, we present an extension of the iterative filtered backprojection that includes a sparsity constraint called soft background subtraction.

This approach is shown to provide sampling artifact reduction when reconstructing sparse objects, and more interestingly, when reconstructing sparse objects over a non-sparse background. The relevance of our approach is evaluated in cone-beam geometry on real clinical data.

1 Introduction

Rotational angiography using C-arm based X-Ray systems provides tomographic acquisitions (scan) of two-dimensional (2D) cone-beam X-Ray projection views, which are used to reconstruct a three-dimensional (3D) model of the human vasculature. To increase the contrast of vessels, a radio-opaque contrast medium must be injected into the blood. Ideally, the injection should be such that all vessels located inside the field of view are fully opacified from the beginning to the end of the contrast scan. The contrast medium being rapidly flushed in the blood flow, the rotation speed of the C-arm must be set as high as safely possible to minimize contrast use, while the detector acquisition frame rate then limits the total number of views, resulting in angular subsampling. Subsampling has little incidence on visualization of highly contrasted structures, but the smaller vessels are affected by streak artifacts originating from intense vessels as well as bone structures. Furthermore, it is common in the clinical practice that opacification

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 97–104, 2011.

occurs later than expected, and thus later than the start of the C-arm rotation, resulting in inconsistently opacified projection views. If the reconstruction is performed on the subset of consistent projections, the vascular structures are then subsampled in a way known as a limited-angle problem.

Depending on the anatomy, it might be advantageous to perform two acquisitions in a single protocol, known as digital subtracted rotational angiography (DSRA), similarly to 2D digital subtracted angiography: a first scan (mask) is performed without contrast injection, while a second scan (contrast) is performed after injection. All background structures (such as bones) are removed by digital log-subtraction of the mask views from the contrast views so that a vessel-only 3D reconstruction can be performed. Because the distribution of vessels can be considered sparse, we propose to address the reconstruction of such subsampled data by using recent compressed sensing results [2,5]. These results have highlighted the interest of casting the reconstruction problem as an optimization problem constrained with a suitable sparse regularization. It has been shown in [2] that perfect reconstruction of a piecewise constant image is achievable through minimization of its total variation (TV). Several works, among which [9] and references within, have focused on the reconstruction of piecewise constant approximations of more clinically relevant type of objects, showing streak artifact reduction, but at the expense of an overall change in the image appearance, that may not be clinically acceptable.

In the case of naturally sparse image such as a vessel tree, the direct use of the image ℓ_1-norm fits the compressed sensing approach without inducing an approximation, and thus is expected to provide streak reduction without changing the nature of clinical images. Still, the resulting iterative procedure for solving the optimization problem may depart from the standard image quality known in the clinical practice. It is indeed a key motivation for using alternative iterative schemes [10]. Nevertheless, in this work, we seek to only address subsampling while deviating as little as possible to the standard filtered backprojection (FBP) type of images. To this aim, we develop an extension of the iterative FBP (iFBP) that includes sparsity constraints. To our knowledge, ℓ_1-regularization of iFBP has never been proposed thus far. A new algorithm, called iFBP with soft background subtraction (SBS), is derived and applied to subsampled subtracted clinical scans. It is shown to provide the expected benefits of streak reduction, in a superior way to what can be achieved applying a positivity constraint. More interestingly, it is further shown that the proposed methodology does also apply to the more general case of sparse structures over a non sparse background, where the positivity constraint is irrelevant.

2 Method

Let $f \in \mathbb{R}^K$ be the vector containing the volume, where K is the number of voxels. Let $p \in \mathbb{R}^I$ be the vector containing the projections, where I is the total number of measurements. Let $R \in \mathbb{R}^{I \times K}$ be the matrix describing the scan trajectory. The reconstruction problem that we address consists in recovering f given the data constraint $p = Rf$.

2.1 Algorithm

A reconstruction method widely used in practice consists of a discretization of the analytical inverse of the system, i.e. backprojection of the filtered projections: $f = R^T Dp$, where R^T is the transpose of R, i.e. backprojection, and D is the ramp filter. Cone-beam dedicated FBP is referred to as the Feldkamp algorithm [6]. Iterative FBP minimizes the Euclidean distance Q between p and the projections of f filtered by $D^{\frac{1}{2}}$ that is expressed in the Fourier domain as the square root of the ramp filter and is such that $D^{\frac{1}{2}} D^{\frac{1}{2}} = D$ [8]:

$$Q(f) = \frac{1}{2} \| D^{\frac{1}{2}} (Rf - p) \|_2^2 \tag{1}$$

iFBP is a fast converging iterative reconstruction algorithm that provides advantages over FBP in handling the cone-beam geometry.

The compressed sensing theory provides a way to recover a signal from a small number of measurements if it is sparse in some basis, that is, if most of the signal energy is concentrated in a few coefficients only. In practice, minimization of the ℓ_1-norm under a data constraint promotes sparsity [5] and can generate sparse approximations of undersampled signals. To get sparse approximations in tomography, we define the functional J as the sum of the previous quadratic data fidelity term Q and a sparsity penalty φ:

$$J(f) = Q(f) + \varphi(f) \tag{2}$$

To minimize J despite the non-differentiablity of φ, we use a simple implementation that consists of an explicit gradient step with step $\tau > 0$ for minimizing Q, which here corresponds to an iFBP iteration, followed by an implicit step for applying constraint φ:

$$\begin{cases} f^{(n+\frac{1}{2})} = f^{(n)} - \tau \nabla Q(f^{(n)}) \\ f^{(n+1)} = \text{prox}_{\tau,\varphi} \left(f^{(n+\frac{1}{2})} \right) \end{cases} \tag{3}$$

The penalization is applied via its proximity operator which is defined as:

$$\forall x \in \mathbb{R}^K, \quad \text{prox}_{\gamma,\varphi} : x \to \underset{y \in \mathbb{R}^K}{\text{argmin}} \left[\gamma \varphi(y) + \frac{1}{2} \|x - y\|_2^2 \right] \tag{4}$$

where $\gamma > 0$ and is set equal to the gradient step τ in (3).

Note that replacing the sparsity constraint φ by the indicator function: $\varphi(f) = \begin{cases} 0 & \text{if } f \in \mathbb{R}_+^K \\ +\infty & \text{otherwise} \end{cases}$, where $\mathbb{R}_+^K$ is the convex set of positive images, is equivalent to taking as proximity operator the projector over $\mathbb{R}_+^K$. Thereby, the scheme presented in (2) also includes the application of the positivity constraint to the iFBP algorithm. We refer the reader to [4] for a complete overview of proximal splitting methods in signal processing.

2.2 Sparsity Constraints

The proximity operator associated to the image ℓ_1-norm: $\varphi(f) = \lambda\|f\|_1$ is the soft-thresholding operator $\mathrm{S}_{\lambda\tau}(f)$ of threshold $\lambda\tau$, where $\lambda \geq 0$ is the regularization parameter that controls the strength of the ℓ_1-penalization. Soft-thresholding is not a filter, but a segmentation step, which does not modify the image appearance above the given threshold.

This sparsity constraint can be combined with the positivity constraint in a single proximity operator that we call soft background subtraction. It is defined by: $SBS_{\lambda\tau}(f) = \begin{cases} f - \lambda\tau \text{ if } f \geq \lambda\tau \\ 0 \quad \text{if } f < \lambda\tau \end{cases}$. It promotes sparsity by removing at each iteration all structures whose intensity is lower than threshold $\lambda\tau$. Considering the case of a sparse image that only contains vessels, there is no strictly positive λ value that may not remove some part of the vessels, and thus biase the reconstructed result.

Consequently, our approach includes defining a set of decreasing λ values: $\Lambda = \{\lambda_n | n = 1, \cdots, N\}$, such that $\lambda_1 \geq \cdots \geq \lambda_N = 0$, where N is the total number of iterations. This strategy is compatible with processing sparse structures over a non-sparse background since the background is reintroduced as λ is decreased.

2.3 Two-Pass Reconstruction

We propose a simple extension of SBS dedicated to the special case of late vessel opacification in the non-subtracted case. To this purpose, we assume that opacification is a binary phenomenon and split the full scan into two subsets: the first subset contains the projection views acquired without opacification, while the second one contains the projection views acquired with opacification. SBS reconstruction is performed on each subset. These two reconstructions taken independently yield limited angular reconstruction of the non-sparse background, but once averaged can recover all background structures.

3 Experiments

We evaluated the reconstruction quality of the proposed algorithm on undersampled acquisitions as found in the clinical practice, where improvement over the standard reconstruction would be welcome. Data are acquired on an Innova 3100 C-arm system (GE Healthcare, Chalfont St. Giles, UK). The system records cone-beam projections at 30 frames/s during an approximately 200 ° rotation at 40 °/s delivering 150 views in total. iFBP algorithm was used with N = 20 iterations and a gradient step $\tau = 0.95$. Direct comparison of iFBP and FBP is difficult because iFBP corrects for cone-beam artifacts and may produce images with slightly higher resolution. Consequently, in the absence of a ground truth, iFBP with positivity constraint is a better reference than FBP, from which establish SBS advantages. For SBS, λ_n was initialized at λ_1 equal to 90 % of the maximum value of the FBP reconstruction and linearly decreased to $\lambda_N = 0$.

For all reconstructions, a volume of 256^3 voxels was computed. Intensities are given in positive Hounsfield Unit (HU), i.e air is 0 instead of -1000 HU.

The first dataset is a DSRA exam of cerebral vessels. A subtracted scan was computed by subtraction of the mask scan from the contrast scan. During the acquisition of the contrast scan, the right vertebral artery (RVA) did not appear opacified during the first half of the rotation. Figure 1 focuses on RVA subtracted reconstruction (note that RVA is the most left vessel in the slices). RVA late opacification resulted in a FBP reconstruction (Fig. 1a) that was deteriorated by horizontal streaks due to the lack of opacification in the lateral views. In addition, the RVA measured intensity (peak: 4540 HU) was half that of the left vertebral artery (LVA, peak: 9929 HU) which was seen fully opacified during the whole scan. The constrained iFBP reconstructions (Fig. 1b for positivity and Fig. 1c for SBS) recovered the RVA from the subset of fully opacified projections (75 views from 100 to 200°) both in terms of shape and intensity (RVA peak: 7541 HU with positivity and 9697 HU with SBS).

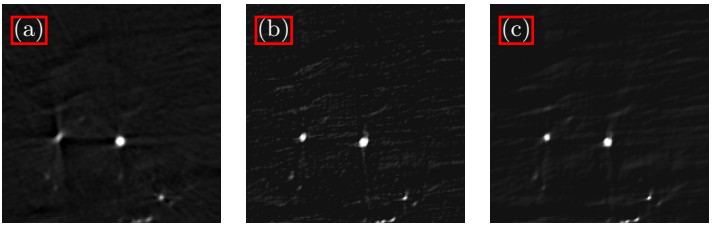


Fig. 1. Subtracted RVA (HU range: -1000 to 6000). (a) FBP reconstruction of 150 views over 0−200 °. (b) iFBP reconstruction of 75 views over 100−200 ° with positivity constraint. (c) iFBP reconstruction of 75 views over 100−200 ° with SBS.

Figure 2 displays a slice higher in the brain, which was not affected by opacification delays and thus reconstructed from the full subtracted scan. Both the positivity constraint (Fig. 2b) and SBS (Fig. 2c) promoted sparsity of the reconstructed structures, thus reducing streak artifacts with respect to the FBP reconstruction of Fig. 2a. However, one sees that the subtracted volume does not contain vessels only, but also perfused tissues (grey areas surrounding the vessels). These areas are not sparse, which shows the importance of an unbiased reconstruction. We computed the mean and standard deviation in a region of interest of 930 voxels within these perfused tissues and found 116 ±142 HU in Fig. 2b for positivity and 96 ±55 HU in Fig. 2c for SBS, a twofold Signal-to-Noise-Ratio (SNR) increase for SBS over positivity (1.8 vs. 0.8), which is confirmed by visual analysis of the respective slices. In terms of resolution, profile curves (not shown) drawn through vessels revealed no differences.

The second dataset is a single scan (no mask) of opacified carotid arteries. Figure 3 compares iFBP with positivity after $N = 5$ and $N = 20$ iterations (Fig. 3a and Fig. 3b, respectively) to iFBP with SBS (Fig. 3c). Removal of the strong horizontal streaks due to a tooth metallic implant could not be obtained after 5

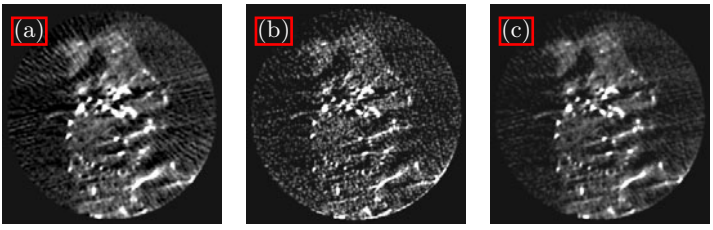


Fig. 2. Subtracted data (HU range: -100 to 500). (a) FBP reconstruction. (b) iFBP reconstruction with positivity constraint. (c) iFBP reconstruction with SBS.

iterations. After 20 iterations, the streaks were removed both by positivity and SBS. Computation of the mean and standard deviation in a volume of interest of 4000 voxels located in the soft tissues between the jaws and posterior to the teeth gave 803 ± 60 HU for Fig. 3a, 787 ± 88 HU for Fig. 3b and 775 ± 49 HU for Fig. 3c. SBS had the highest SNR (15.8 vs. 13.4 and 8.9), despite the 20 iterations. We attribute this SNR improvement to the reduction of intensity of smaller streaks in the soft tissues. Profile curves (not shown) through the bony structures of Fig. 3b indicated a possibly slightly higher resolution for the positivity constraint after 20 iterations, although possibly linked to the strong noise increase. On the other hand, peak values over the profiles were not systematically higher in Fig. 3b than in Fig. 3c.

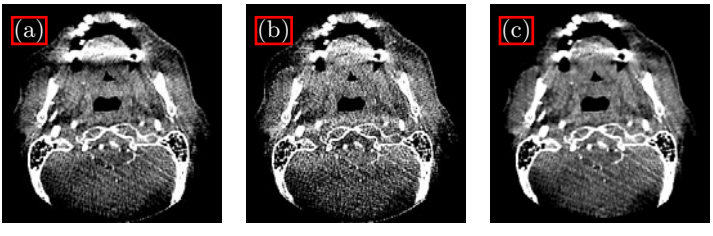


Fig. 3. Second dataset (HU range: 550 to 1050). (a) iFBP reconstruction with positivity constraint – 5 iterations; (b) iFBP reconstruction with positivity constraint – 20 iterations; (c) iFBP reconstruction with SBS – 20 iterations.

Application of the two-pass reconstruction is shown in Fig. 4. The same slice as in Fig. 1 was reconstructed without subtraction. Besides what we already noted in the subtracted case (Fig. 1a), the non-subtracted FBP reconstruction (Fig. 4a) showed that the horizontal and vertical black streaks associated to the RVA hid the underlying tissues. Limitating FBP reconstruction to the second half of the scan (Fig. 4b) yielded no visual improvement of the RVA while degrading both LVA and background. On the contrary the SBS constraint (Fig. 4c) recovered both arteries, as in the subtracted case. Finally, the two-pass reconstruction was necessary to also recover the background as shown in Fig. 4d. However, the averaging operation of the two-pass reconstruction affected the quantification of

the RVA: RVA peak value was 11744 HU for Fig. 4c and 6353 HU for Fig. 4d (similar to the RVA peak value of 6265 HU for FBP). The mean and standard deviation were computed in a volume of interest of 40000 voxels located in the uniform tissues between the two arteries, yielding 1286 ±318 HU for FBP (Fig. 4a) and 1070 ±215 HU for the two-pass algorithm (Fig. 4d), a 25% SNR increase for the two-pass algorithm over FBP due to streak reduction.

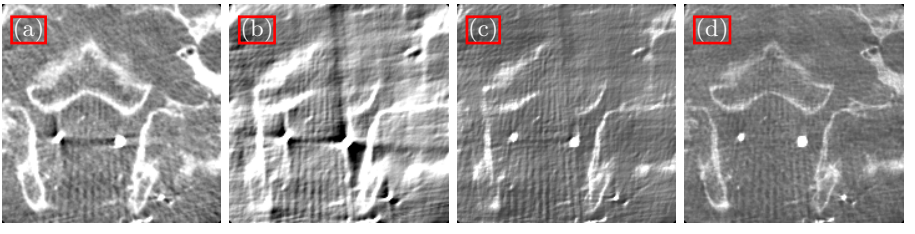


Fig. 4. Non subtracted RVA (HU range: 0 to 2400). (a) FBP reconstruction of 150 views over 0−200 °. (b) FBP reconstruction of 75 views over 100−200 °. (c) iFBP reconstruction of 75 views over 100−200 ° with SBS. (d) Two-pass reconstruction.

4 Discussion and Conclusion

In this study, a new ℓ_1-regularization of the iterative filtered backprojection, called soft background subtraction, has been introduced and applied to a selection of clinical datasets that suggests a domain of applicability: clinical cases of limited frame rate or of limited angular coverage. It has been shown to mitigate or reduce streak artifacts, when reconstructing sparse volumes from sparse data. Because we voluntarily did not model noise or other physical aspects of the imaging device beyond the cone-beam geometry, the reconstructed images had a similar appearance than FBP-type reconstructions.

The positivity constraint was effective at reducing streaks in subtracted images. Indeed, it is a special case of the proposed background subtraction technique, that can still be considered a sparse regularizer since negative voxels that get set to zero are many in this context. More interestingly, the linear decrease of the soft background subtraction threshold provided superior reconstruction of the non-sparse background of sparse high-intensity structures, resulting in more uniform tissue depiction when affected by streaks from either teeth, bones or opacified vessels, at equal resolution. Linearly decreasing the regularization strength is too simple a heuristic to exactly solve the hard-constrained compressed sensing problem, and is thus surely suboptimal. Still it has an intuitive interpretation that makes it well suited for a clinical use. Most importantly, the generated solution is not a compromise between a fitting term and a penalty, but targets an unbiased fit to the data. SBS provided to a certain extent the expected benefits of compressed sensing theory: reduced streaks. Such gains would otherwise be achievable only by increasing the number of projections, which was not possible in these clinical settings, rather than increasing the X-Ray dose of each projection image.

The soft background subtraction approach is applicable to any tomographic dataset to provide localized and specific improvements regarding subsampling while appearing just as good as FBP otherwise. The combination of iterative FBP and simple thresholding operations entails a computation time of roughly twice the standard FBP computation per iteration. Further studies are required to determine whether the number of iterations can be reduced.

The two-pass algorithm recovered the shape of the degraded vessel with its background, but not its quantification. To recover both, we need to recast the problem as a dynamic one. Compressed sensing based tomographic reconstruction of dynamic series is an active research area [3,7] that implies to address higher levels of undersampling than those currently seen in the clinical practice. We anticipate that the proposed approach will provide a key brick to such reconstructions, as proximal splitting allows for combining iterative FBP with a variety of non-linear filters or segmentations, that can be derived by the calculus of variations [1].

References

1. Aubert, G., Kornprobst, P.: Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations. Applied Mathematical Sciences, vol. 147. Springer, Heidelberg (2006)
2. Candès, E., Romberg, J., Tao, T.: Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information. IEEE Transactions on Information Theory 52(2), 489–509 (2006)
3. Chen, G.H., Tang, L.: Prior Image Constrained Compressed Sensing (PICCS). Medical Physics 35(2), 660–663 (2008)
4. Combettes, P.L., Pesquet, J.C.: Proximal Splitting Methods in Signal Processing. In: Bauschke, H.H., Burachik, R.S., Combettes, P.L., Elser, V., Luke, D.R., Wolkowicz, H. (eds.) Fixed-Point Algorithms for Inverse Problems in Science and Engineering, Springer Optimization and Its Applications., vol. 49, ch.10, pp. 185–212. Springer, Heidelberg (2011)
5. Donoho, D.L.: For Most Large Underdetermined Systems of Linear Equations, the Minimal ℓ_1 Solution is Also the Sparsest Solution. Communications on Pure and Applied Mathematics 59(6), 797–829 (2006)
6. Feldkamp, L.A., Davis, L.C., Kress, J.W.: Practical Cone-Beam Algorithm. Journal of the Optical Society of America A 1(6), 612–619 (1984)
7. Jia, X., Lou, Y., Dong, B., Tian, Z., Jiang, S.: 4D Computed Tomography Reconstruction from Few-Projection Data via Temporal Non-local Regularization. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6361, pp. 143–150. Springer, Heidelberg (2010)
8. Riddell, C., Savi, A., Gilardi, M., Fazio, F.: Frequency Weighted Least Squares Reconstruction of Truncated Transmission SPECT Data. IEEE Transactions on Nuclear Science 43(4), 2292–2298 (1996)
9. Ritschl, L., Bergner, F., Fleischmann, C., Kachelrieß, M.: Improved Total Variation-Based CT Image Reconstruction Applied to Clinical Data. Physics in Medicine and Biology 56(6), 1545–1561 (2011)
10. Thibault, J., Sauer, K., Bouman, C., Hsieh, J.: A Three-dimensional Statistical Approach to Improved Image Quality for Multi-slice Helical CT. Medical Physics 34(11), 4526–4544 (2007)

Real-Time 4D Ultrasound Mosaicing and Visualization

Laura J. Brattain[1,2] and Robert D. Howe[1]

[1] Harvard School of Engineering and Applied Sciences, Cambrige, MA 02138
[2] MIT Lincoln Laboratory, 244 Wood St., Lexington, MA 02420
{brattain,howe}@seas.harvard.edu

Abstract. Intra-cardiac 3D ultrasound imaging has enabled new minimally invasive procedures. Its narrow field of view, however, limits its efficacy in guiding beating heart procedures where geometrically complex and spatially extended moving anatomic structures are often involved. In this paper, we present a system that performs electrocardiograph gated 4D mosaicing and visualization of 3DUS volumes. Real-time operation is enabled by GPU implementation. The method is validated on phantom and porcine heart data.

Keywords: 3D ultrasound, electromagnetic tracking, graphic processing unit, volume mosaicing, volume registration.

1 Introduction

Beating heart procedures eliminate the need for cardiopulmonary bypass and its associated morbidities [1], and allow the surgeon to evaluate the procedure under physiologic loading conditions. Real-time 3D ultrasound (3DUS) imaging is often the imaging modality of choice for guiding these procedures. 3DUS is non-invasive, inexpensive, able to see through the blood, and can provide real-time diagnostic information of the anatomical region of interest. 3DUS also mitigates the difficulties in spatial perception associated with traditional 2DUS [2].

Intra-cardiac procedures often involve geometrically complex and spatially extended moving anatomic structures. Real-time visual feedback of the moving structures is critical to the success of these procedures. In order to realize the potential of 3DUS in guiding such procedures, two challenges must be addressed: high noise and narrow field of view. These limitations can be mitigated with a real-time panoramic view of the region of interest.

Electrocardiograph (ECG) gating is commonly used in the 4D reconstruction of a time-series of 3DUS volumes. Efforts have been reported in the reconstruction of gated 4DUS by mosaicing tracked 2DUS imaging [3][4][5]. In this case, 2DUS frames are acquired at the same point in the motion cycle of the target and composited into 3D volumes first, which then are assembled into a time series for sequential display.

With the arrival of real-time 3DUS machines, 3DUS volumes can be obtained directly and they offer improved spatial information compared to 2DUS, especially for geometrically complex targets with rapid motion. Brekke *et al* developed an algorithm that assembles ECG gated 3D cardiac ultrasound sub volumes into a symmetric pyramidal without using image registration [6]. They reported decreased

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 105–112, 2011.

geometric distortions compared to the non-stitched 3D volume. One major limitation of such a system is that it requires acquiring sub volumes at specific azimuth and elevation angles, and as a result, users have limited control over probe placement and orientation. Kutter *et al* presented a multi-modal 3DUS registration and mosaicing approach by incorporating information from co-registered CT [7]. However, CT data is not always available for clinical cases, especially for the non-intervention cases. In addition, pre-operative CT data usually does not accurately represent the real-time state of a patient's condition.

In this paper, we present the first system that performs freehand real-time 4DUS reconstruction and visualization based on streaming 3DUS data and ECG gating. The freehand capability is important as it allows clinicians to optimize the ultrasound acquisition geometry. Volume registration is performed through tracking the 3DUS probe using an electromagnetic (EM) tracking system. This method is similar to the reconstruction of 2D images as previously described, with a key difference being that the data throughput is one or two orders of magnitudes higher in 3D imaging. We address this issue by leveraging the parallel computing power of graphic cards. The computation of the volume mosaicing and rendering are done using a parallel implementation. In this paper, we begin with a description of the system setup, and then present our 4D reconstruction process, followed by results and discussions.

2 System Configuration

The system consists of three major hardware components: a GPU enabled computer (Dell Alienware Aurora, Intel Core i7 processor at 2.67GHz, 6GB RAM, NVIDIA GTX260 graphics cards), a 3DUS scanner (Philips SONOS 7500 with x4 probe imaging at 8cm depth and 28Hz) that can stream 3DUS images to the PC, and an EM tracking system (trackStar 3D Guidance System, Ascension Technology). The EM sensor is rigidly mounted to the 3DUS probe, and through a calibration process, we obtain the transformation between the ultrasound volume and the EM coordinates.

2.1 Mosaicing Process

The software application is implemented in C/C++, using multi-threading for data acquisition, CUDA (driver v.3.1.3.) for parallel computing, and OpenGL for visualization. At run time, in addition to the main thread, there are three threads running in parallel on the CPU: 3DUS volume streaming, EM tracking streaming, ECG waveform streaming. ECG phase time stamping is used for the acquisition of 3DUS volumes, which in turn are sent to the GPU, along with the EM tracking coordinates. The computations of volume registration, mosaicing and ray-casting are performed on the GPU. In order to perform mosaicing, we first calculate the transformation matrix between the input volume and the reference volume. The transformation matrix is determined by the EM tracker's outputs and the calibration matrix between the 3DUS coordinate system and the EM tracker's coordinate system. Fig. 1 illustrates the data flow between CPU and GPU.

Since the volume registration and compositing is a highly parallelizable process, we developed CUDA kernels to facilitate the computation. On the GPU, the transformation matrix is mapped to constant memory, and both the input and

reference 3DUS data are mapped to CUDA 3D textures. Image registration uses tri-linear interpolation, after which the registered volume and the reference volume are composited and the 3D texture that holds the reference frame is updated to the composite volume. There are a number of ways for compositing; we currently use the maximum intensity at the overlapping locations, because it eliminates the need to mask the void in the composite volume. To minimize computationally expensive memory transfer between host and device, the composite volume is sent to the ray-casting kernel directly through device-to-device memory copy for rendering.

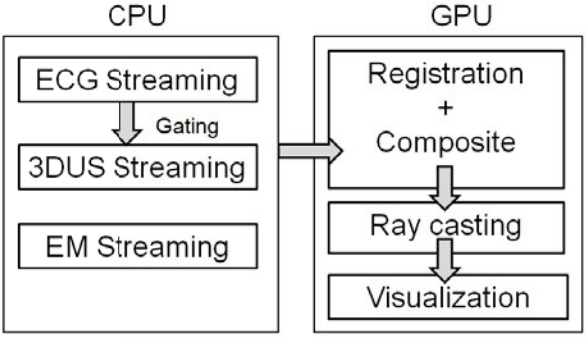


Fig. 1. Data flow between CPU and GPU

2.2 ECG Gating

Cardiac synchronization limits artifacts due to motion of the heart, thus enabling clear visualization across the phases of the cardiac cycle. The ECG R wave peak is typically used for this purpose as it can be readily identified in the ECG signal. The interval between R wave peaks, the RR interval, defines one cardiac cycle for gating. In our study, we perform R wave peak detection in real-time and divide an RR interval into a user-specified number N of equally spaced bins.

The user first chooses a location to image and sets it as the reference frame. N bins of gated data are collected at this location and saved as the reference volumes. The subsequent volumes acquired are added to the final mosaiced volume by being registered into the corresponding reference volume with the matching bin number, as above. To acquire these volumes, the user moves the probe to a desired location and triggers the data acquisition at that location. The mosaicing is performed immediately and can be visualized at real time. The final time-series consists of N mosaiced volumes. Fig. 2 illustrates this process.

3 Experimental Validation

3.1 Methods and Results

We use phantom data in a water tank to develop and test the system. The phantom consists of horizontal, vertical and diagonal strings suspended across a rigid frame

with a dimension of 8 x 6 cm (Fig. 3A). A string phantom provides an easy means to validate the mosaicing result by confirming the mosaiced colinearity of the segments from each input volume. The phantom is suspended in the water tank with one side connected rigidly to a cam device. The cam device simulates the cyclic motion of the mitral annulus (Fig. 3B). A signal from the cam device simulates the output of an ECG monitor. Fig. 4 (A) shows one mosaiced result of the string phantom, and (B) – (E) are four of the 3DUS volumes used in the creation of the mosaiced volume. It is evident that the mosaiced volume provides a wider field of view. The input volume has a dimension of 128 x 48 x 208 voxels, and the dimension of the mosaiced volume is 320 x 144 x 210 voxels. On average, the time used to perform every two volume registration and stitching on the GPU is 25 ms on the system described above. The time to mosaic 8 bins is 175 ms. This mosaicing process is applied to each bin of the motion cycle (Fig. 5).

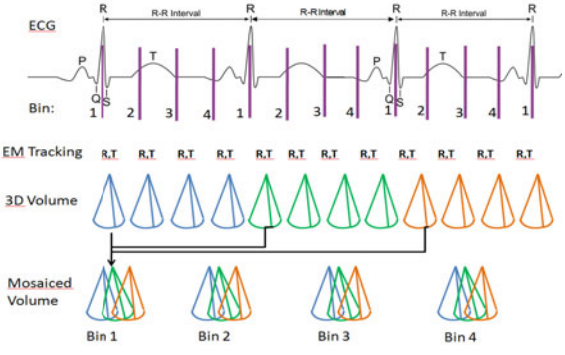


Fig. 2. 4DUS data mosaicing with ECG gating

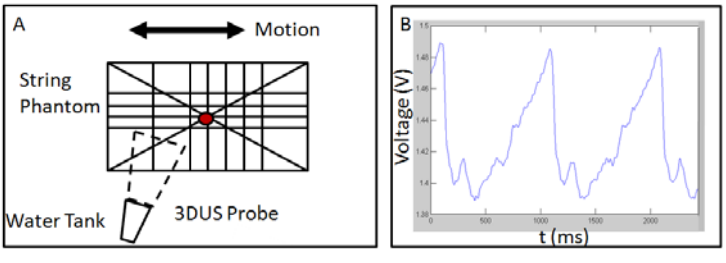


Fig. 3. Experiment setup. (A) Water tank with string phantom. The red dot indicates a fiducial point in subsequent figures. (B) Periodic motion trajectory.

We then conducted experiments using a porcine heart in a water tank. As with the string phantom, the heart is mounted on a plastic frame which is rigidly connected to the cam device. The heart (Fig. 6) is positioned sideways, with the top and bottom portions sectioned and removed, allowing water to flow through the heart, which

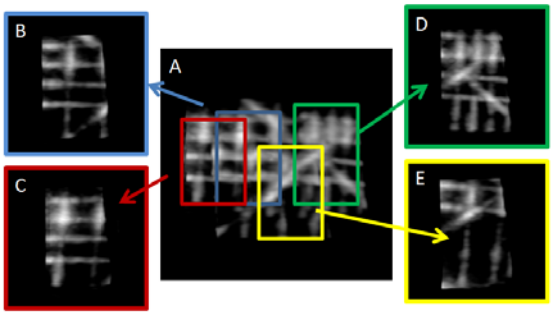


Fig. 4. (A) One mosaiced result. (B)-(E) Four of the input volumes.

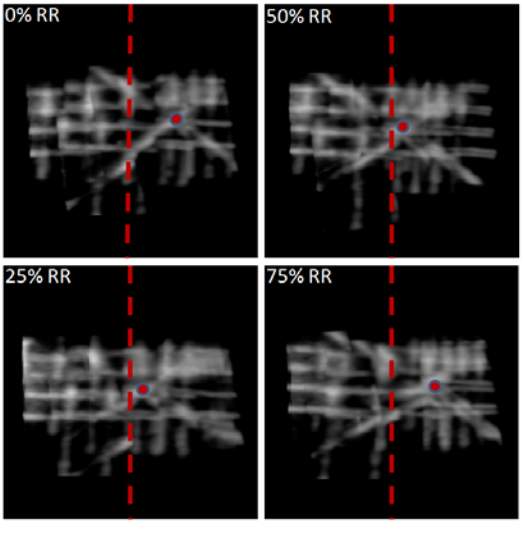


Fig. 5. Mosaiced results at different RR intervals. The red dot is the fiducial point on the string phantom. The red dotted line indicates the mid line of the cross section. The varying distance between the fiducial point and the mid line is the result of the motion of the phantom.

causes the valves to open and close periodically. The ultrasound probe is positioned adjacent to the heart at the region where the right atrium and tricuspid valve are located (Fig. 6 (D)).

Fig. 6 (E) is the result of mosaicing 18 3DUS volumes at each of N=8 time bins, using a total of 144 volumes. The red arrow indicates the tricuspid valve, which is open at 0% RR interval and closed at 50% RR interval. This periodic motion can be seen clearly in the visualization loop of the complete time series. This highlights the value of the visual feedback for real-time intra-cardiac procedural guidance.

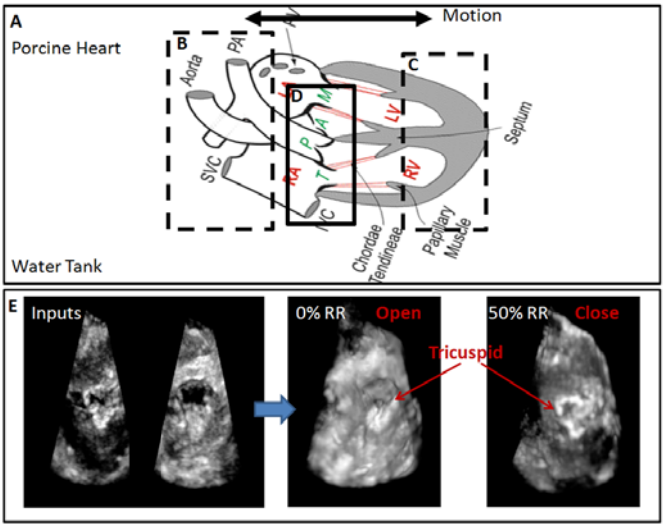


Fig. 6. (A) Porcine heart [8]. The sectioned portion B and C are removed. (D) Imaging section. (E) Two Input volumes (left) Mosaicing results at 0% and 50% RR interval (right).

3.2 Validation

We use the string phantom to validate the co-linearity of the mosaiced line segments. For a mosaiced line, we first identity the line segment from the reference frame since it is the segment that the other segments are registered to. This line segment is used as the centroid of the mosaiced line. We then calculate the number of non-trivial voxels around the centroid at a given radius. We expect to see that most of the contributing voxels are around the centroid, and the number of contributing voxels falls off quickly when the radius increases. Fig. 7 is a plot of the percentage of the contributing voxels vs. radius from the centroid. The width of the mosaiced line is roughly 4 voxels, which equals the width of the line segment in the reference frame.

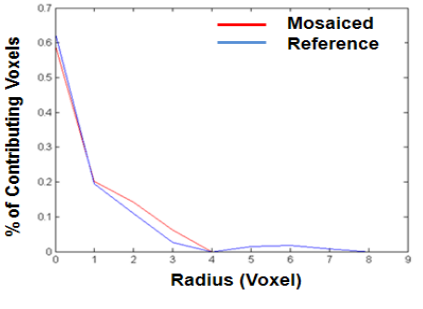


Fig. 7. Co-linearity of mosaiced line segments

4 Discussion

We have developed the first real-time freehand 4D mosaicing and visualization system using ECG gated 3DUS streaming data. 3DUS streaming provides a much higher data rate than 2DUS based 3D reconstruction. This allows faster updates of the mosaiced image, but requires parallel computation. Although present 3DUS machines have a lower spatial resolution than the best 2DUS machines, 3DUS resolution is improving rapidly. For example, newer machines (e.g. the Philips iE33 scanner) have 2-3 times the spatial resolution of the machine used in this study.

4.1 Temporal Resolution and Imaging Time

There is a tradeoff between imaging time to build or update the mosaic and time resolution (i.e. the number of time bins N). Temporal resolution is particularly important for accurately capturing the motion of fast-moving structures such as cardiac valves. For example, the peak velocity of the mitral valve annulus has been estimated as $v = 210$ mm/s [9], so assuming a typical RR interval of $T_{RR} = 1$ s and with number of bins $N = 10$, the spatial blur within one time bin can be as high as $dx = v\,(T_{RR}/N) = 21$ mm. This shows that high temporal resolution reduces the effect of spatial blurring.

On the other hand, higher temporal resolution will result in longer data acquisition and processing time and higher demand on computer memory size. In addition, the current 3DUS streaming is at approximately 30 Hz, which limits the data acquisition to 30 volumes per second.

4.2 Real-Time Overlay

One of the important features that this system offers is a “big picture” view of a large section of the heart, combined with a “fovea” real-time view of the smaller region where the procedure is performed. The mosaiced time series can be viewed as a map in the background, which can be extremely useful to the user for navigating to a specific location and for planning subsequent surgical steps. The specific location at which the catheter or instrument is working is then automatically overlayed on the mosaiced times series in real-time. Any region of the mosaic can be updated by simply moving the probe to view that region, and the real-time view follows the instrumentation. Color coding of real-time versus previously mosaiced data can inform the clinician of which regions represent the current versus historical images. This allows for simple and reliable control of the image acquisition and display process.

5 Conclusion

In this paper, we presented a system for real-time 4DUS mosaicing and visualization. The system integrates EM tracking systems and GPU implementation for real-time registration and mosaicing of high data rate 3DUS images. ECG time stamping provides synchronization with heart motion. Temporal resolution can be controlled by

the user, and real-time data is continuously overlayed on the mosaic. The system merges previously-acquired wide-area data with real-time focal data, facilitating navigation and procedure execution with a simple user interface.

Acknowledgments. The Harvard University portion of the work is sponsored by US National Institutes of Health under grant NIH R01 HL073647-01. The MIT Lincoln Laboratory portion of the work is sponsored by the Department of the Air Force under Air Force contract #FA8721-05-C-0002. Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the United States Government.

References

1. Murkin, J.M., Boyd, W.D., Ganapathy, S., Adams, S.J., Peterson, R.C.: Beating Heart Surgery: Why Expect Less Central Nervous System Morbidity? Annals of Thoracic Surgery 68, 1498–1501 (1999)
2. Cannon, J.W., Stoll, J.A., Salgo, I.S., Knowles, H.B., Howe, R.D., Dupont, P.E., Marx, G.R., del Nido, P.J.: Real-time three dimensional ultrasound for guiding surgical tasks. Computer Aided Surgery 8(2), 82–90 (2003)
3. Treece, G.M., Prager, R.W., Gee, A.H., Cash, C.J.C., Berman, L.: Grey-scale gating for freehand 3D ultrasound. In: IEEE. Int. Symp. on Biomedical Imaging (2002)
4. Pieper, S., Halle, M., Kikinis, R.: 3D Slicer. In: IEEE Int. Symp. on Biomedical Imaging (2004)
5. Pace, D., Gobbi, D., Wedlake, C., Gumprecht, J., Boisvert, J., Tokuda, J., Hata, N., Peters, T.: An open-source real-time ultrasound reconstruction system for four-dimensional imaging of moving organs. In: Systems and Architectures for Computer Assisted Interventions. MICCAI 2008 Workshop (2008)
6. Brekke, S., Rabben, S.I., Stoylen, A., Haugen, A., Haugen, G.U., Steen, E.N., Torp, H.: Volume Stitching in Three-Dimensional Echocardiography: Distortion Analysis and Extension to Real Time. Ultrasound in Medicine and Biology 33, 782–796 (2007)
7. Kutter, O., Wein, W., Navab, N.: Multi-modal Registration Based Ultrasound Mosaicing. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5761, pp. 763–770. Springer, Heidelberg (2009)
8. CV Physiology, `http://www.cvphysiology.com/Heart%20Disease/HD001.htm`
9. Kettler, D.T., Plowes, R.D., Novotny, P.M., Vasilyev, N.V., del Nido, P.J., Howe, R.D.: An Active Motion Compensation Instrument for Beating Heart Mitral Valve Surgery. In: IEEE Intl. Conf. Intelligent Robots & Systems (2007)

Robotic Hair Harvesting System: A New Proposal

Xiang Lin, Toji Nakazawa, Ryuya Yasuda, Etsuko Kobayashi, Ichiro Sakuma, and Hongen Liao

Graduate School of Engineering, University of Tokyo, Japan
shanelinnz@gmail.com, liao@bmpe.t.u-tokyo.ac.jp

Abstract. Follicular Unit Extraction (FUE) has become a popular hair transplanting method for solving male-pattern baldness problem. Manually harvesting hairs one by one, however, is a tedious and time-consuming job to doctors. We design an accurate hair harvesting robot with a novel and efficient end-effector which consists of one digital microscope and a punch device. The microscope is first employed to automatically localize target hairs and then guides the punch device for harvesting after shifting. The end-effector shows average bias and precision of 0.014 mm by virtue of a rotary guidance design for the motorized shifting mechanism.

1 Introduction

At least half of the male population by age 50 in this world have a hair loss problem. About 250,000 hair restoration procedures were estimated to be performed worldwide in 2008 [1]. Hair transplanting surgery is considered as the most efficient solution to this problem.

The follicular unit extraction (FUE) technique, proposed in [3], is considered as one of the most advanced methods for transplanting. It uses a small sharp punch to directly extract follicular units one by one from man's scalp to avoid leaving a linear scar. The linear scar is often observed in the currently most popular method, follicular unit transplanting (FUT) [2]. FUE has been proven less invasive, faster recovery, and less noticeable scarring. The main disadvantages are related to potentially high rates of follicle transection and the requirement of a high degree of technical skill to doctors, which prevents the technique from being widely accepted.

A robotic system can assist doctors to both improve the accuracy and reduce the variability in operation results due to an individual's technique. Little research has been done in this area in last decade. Onda et. al. claimed a robotic harvesting system in [4] but it still had to be manually held by doctors. Gildenberg proposed a real robotic system in [5], which used a stereotactic video system for 3D reconstruction of target hairs. The system was further implemented in [6] but unfortunately no detail report has been presented to demonstrate how accurate the system is by using the stereotactic system.

In this paper, we design a new accurate harvesting robot with a novel end-effector which consists of one digital microscope (a single camera) and a punch

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 113–120, 2011.

device. The microscope is employed as the first step in the process to automatically localize target hairs with a customized image-based visual servoing method. The punch device is then shifted by a motor to the same position as the microscope for harvesting. R-guide is applied to provide the shifting mechanism which greatly improves the system accuracy and simplifies the mechanical design. In this paper we focus on the system design, particularly the end-effector design.

The remainder of the paper is organized as follows. The design of the whole system is described in Sect. 2. Section 3 gives the details of the end-effector. The calibration method for the end-effector is provided in Sect. 4. Experiments and results are presented in Sect. 5, followed by conclusions in Sect. 6.

2 System Design

The target of the system is to accurately extract hairs one by one from a donor area. The harvesting process of a single hair includes the following steps. Firstly, the system finds a suitable hair in the field of view and move the punch device in front of it. A sharp needle is then inserted into the skin along the hair axis and dissects the whole hair follicle from the surrounding tissues. The insertion terminates once the follicle is extracted or a fixed depth is met. For safety reasons, the system cannot be designed as fully automatic and the whole process must be under doctor's supervision.

Figure 1 shows the design of our proposed system which consists of four main parts, a robot arm, an end-effector, a computer, and a user interface. The robot arm (VP-6242E, DENSO, Japan) has six degrees of freedoms (DOFs) and is able to move to any pose with the repeatability of 0.02mm within the region of interest. The end-effector includes a punch device for harvesting and a digital microscope for inspection and localization. The two devices switch about each other to complete both the localization task and the harvesting task. The user interface has three functions, monitoring the process in real time, assisting doctors to choose suitable hair candidates, and receiving commands from doctors. The computer is used for image processing, image acquisition and robot arm controlling. The process has the following two phases:

Localization phase

1. The digital microscope is shifted by a stepping motor to the working position shown in Fig. 2A. Doctors select which hair should be harvested by clicking on the image captured from the microscope.
2. An image processing program is executed to obtain the hair's orientation and position on the 2D image.
3. The orientation and position information is fed to a customized visual servoing software. The software calculate the necessary movement that the robot arm needs to move the hair to a target pose.
4. The movement commands are sent to the robot arm for action and the image on the screen is updated after moving.

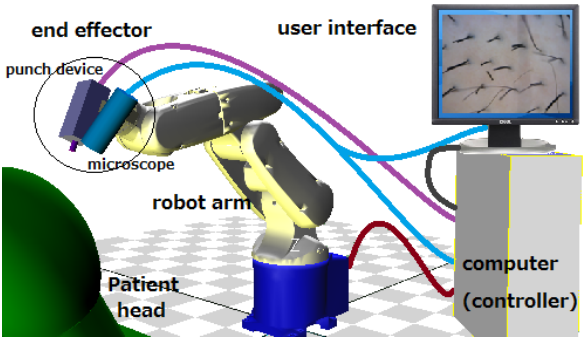


Fig. 1. Hair harvesting system, consisting of a robot arm, an end-effector, a user interface and a computer

5. The image processing program tracks the position and orientation of the hair in the updated image.
6. The robot arm stops once the hair is presented with the target pose in image. Otherwise a recursive process from Step 3 to Step 5 is executed until the hair reaches the target pose.

It should be noticed that different target poses may be available for use but a suitable pose should have a unique presentation in the image and must have enough information for the image-based visual servoing to process. We proposed of using the pose which is located on the microscope optical axis and have the same orientation as the axis. The hair should be shown as a black dot at the center of the image (Fig. 3E) once the pose is reached. The distance between the microscope and the target hair can be determined by the analysis of image blurriness because the microscope with high magnification often has a very short depth of field.

Harvesting phase

1. The punch device is shifted to the working position (Fig. 2B). The needle of the punch device is advanced into the skin by a linear actuator.
2. The needle stops once the embedded sensor detects that the hair has already been dissected from the surrounding tissues or a maximal distance under the skin is reached.
3. The needle is withdrawn from the skin and the current process finishes. Doctors determine if a new process will be started by selecting another hair on the screen.

There is no visual servoing in the harvesting phase. The system is simply designed in a open-loop way without active compensation for misalignment from the shifting. The accuracy is relied on mechanical calibration which has been proven to meet the requirements in Sect. 4.

3 End-Effector Design

A reliable and cost-efficient mechanical design for the shifting mechanism is critical to the system accuracy. In this paper, we use the R-guide (HCR 15A+60/150R, THK, Japan) for high-accurate rotary guidance in a large working area. The microscope and the punch device are installed on the R-guide through an arc-shaped rack. A small angle less than 30 degrees is set between them to reduce the shifting time. The motion of the R-guide is driven by a light weighted stepping motor (CMK223AP-SG7.2, Oriental Motor, Japan) through the gear at the rim of the rack. There are two main advantages of using a R-guide here:

1. Able to reduce system errors: The working area is designed at the center of the R-guide. Any position error caused at the rim of the rack can be reduced at the center. For example, there often exists a backlash between the pinion and the gear on the rack. If the backlash is assumed to be 1.0 mm and the radius of the rack is set to 160 mm, the angular error caused by the backlash will be about 0.36 degrees. In an ideal situation, if the needle exactly passes the center and a hair is just located at the center, this angular error will cause about 0.034 mm position error at the tip of a 5.5 mm long hair (the average hair length after shaving in transplanting surgeries). This capability will also be proved by the repeatability tests in the experimental section.
2. Able to monitor needle insertion with microscope: It is necessary to monitor the depth of insertion in real time for safety purposes. One solution is to measure the needle length which is left above the skin. It can estimated from the image captured by the microscope. The requirement is that the distance between the working space and the microscope must be kept constant since the microscope has a short depth of field. This can be satisfied in this design. The insertion area above the skin can always be clearly viewed from the microscope during the whole process. Some specific marks are needed on the needle for image analysis.

The needle insertion is served by a linear actuator/motor (CCM05M-072-C-D01-EN-E1, THK, Japan) which has a high resolution of 1.64 μm and a relative high speed of 1.0 m/s. The actuator thrust can go up to 10.4N which is enough to pierce human's scalps. It can be automatically deactuated to provide a safe stop function when the maximum thrust lasts more than several seconds. The needle has an outer diameter of 0.9 mm and a 0.1 mm thick wall. It is installed at the tip of a hollow spline shaft and controlled by the linear actuator. In order to automatic extract the follicular unit after harvesting, a vacuum system is connected to the other end of the shaft. A photoelectric sensor (not shown) will be installed on the path to the vacuum flow. It stops the linear actuator from further insertion once the hair is detected.

The microscope is attached to the rack by an adjustable bracket which can be manually adjusted in 3 DOFs for calibration. Detail adjustment will be described in Sect. 4. A high speed camera of 120 fps and 659 x 494 resolution (Prosilica GC660, AVT, USA) is applied in this application. The lens is adjustable with the magnification from 5x to 50x.

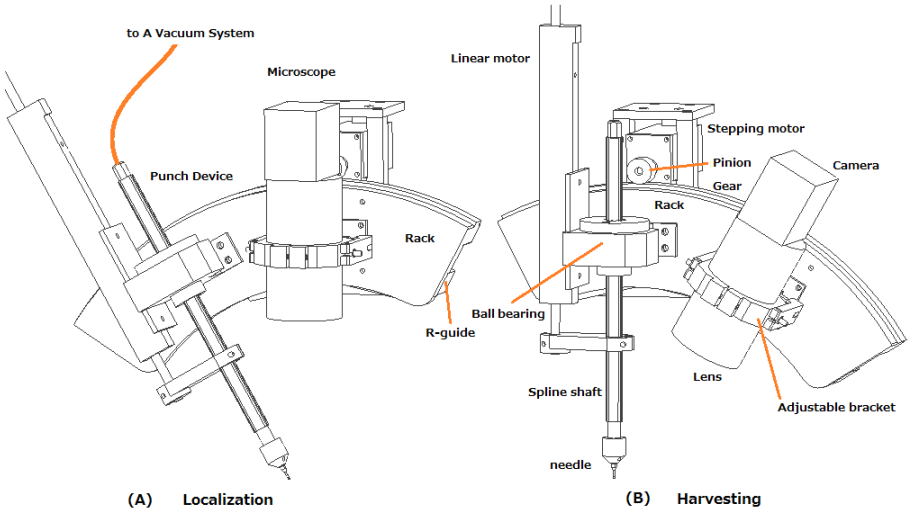


Fig. 2. Two phases, (A) localization phase when the microscope is in the working position, (B) harvesting phase when the punch device is in the working position

4 End-Effector Calibration

The system could provide high accuracy only if the needle insertion axis was exactly located as same as the microscope axis when they were in the working position. A new calibration method of the end-effector was proposed here for this alignment. It was actually a hair implanting process, the reverse of a normal harvesting process. The hair phantom used here was a small piece of 0.5 mm size pencil lead. It was implanted into a flat stage in the same orientation as the shaft axis and also on the path where the needle passed. The microscope position was then adjusted until its axis had the same direction as the pencil lead. The details are described as follows.

1. Implant the pencil lead into the stage with a specific pose: Some silicon gel was put on the stage in advance. The punch device was shifted to the working position (Fig. 3A). The pencil lead was manually inserted into the needle. It was slowly advanced to the stage (Fig. 3B) with the needle and inserted into the gel on the stage. The lead was left in the gel until the gel became dry. In this way, it could be ensured that the lead had the same orientation as the shaft and was also located on the path of the needle.
2. Adjust microscope's orientation and position: The microscope was now shifted to the working position. The lead might not be located at the image center (Fig. 3D) or even outside of the field of view. The 3 DOFs bracket (Fig. 3C), with two rotations and one translation, was used for the adjustment. The translation in the horizontal direction(green arrow) would be made up by the rotation of the stepping motor. It should be noticed that Rotation A and Translation C must be adjusted together since both of them would change

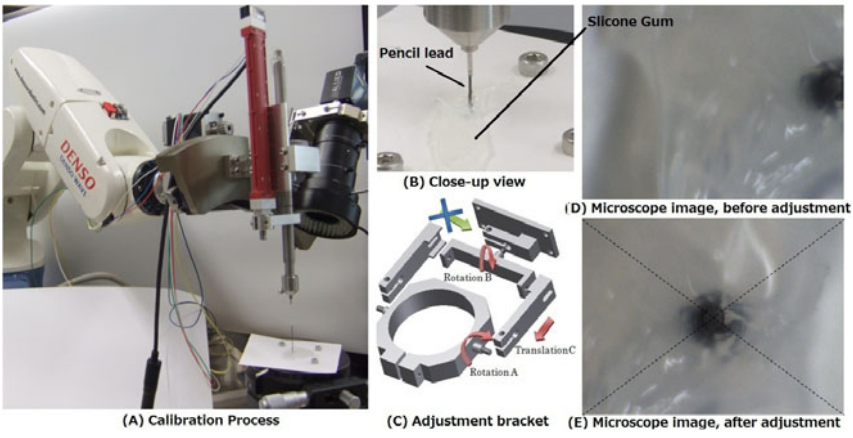


Fig. 3. Calibration process. (A) the stage and two steps in the process, (B) a close-up view of the fixing the pencil lead by silicone gel, (C) 3 DOFs adjustable bracket, (D) before adjustment, and (E) shown as a black dot at the center after adjustment

the vertical position, while Rotation A would change the orientation as well. After adjustment the lead should be shown at the image center (Fig. 3E).

5 Experiments and Results

All the evaluations were performed on 2D images obtained from the microscope. Microscope pose repeatability, position errors and orientation errors were evaluated. A piece of paper with a number of 3.0 mm size black circles was fixed on the adjustable stage which had been used in the calibration step (Fig. 4A). The circle centers were clearly marked with crosses in Fig. 4D.

Pose Repeatability: An image such as Fig. 4D was captured from the microscope. The repeatability was evaluated from the difference between this image and subsequent images repeatedly taken after fixed number of shifts. In this experiment one image was taken after every 5 times of shifts and 10 images were captured in total. On each image the boundary of the circle was first manually marked by green dots (Fig. 4E), followed by a circle fitting to automatically detect the centroid. The repeatability error was -1.2 $\pm$ 0.5 μm in x direction and -0.1 $\pm$ 0.3 μm in y direction in mean $\pm$ std format.

Position errors in x and y directions: A similar process was performed for the computation of position errors. The microscope was set to the working position. Instead of moving the microscope, the stage was translated within its plane to drive a circle in the image to the image center (Fig. 4D). The punch device was then shifted to the working position and make a hole on the paper. The hole position was recorded by the microscope when shifted back. Both the boundary of the black circle and the hole were manually marked (Fig. 4E). Circle fittings were performed to obtain both centroids and radii. The distance between the hole centroid and the image center were measured as the position errors. In order

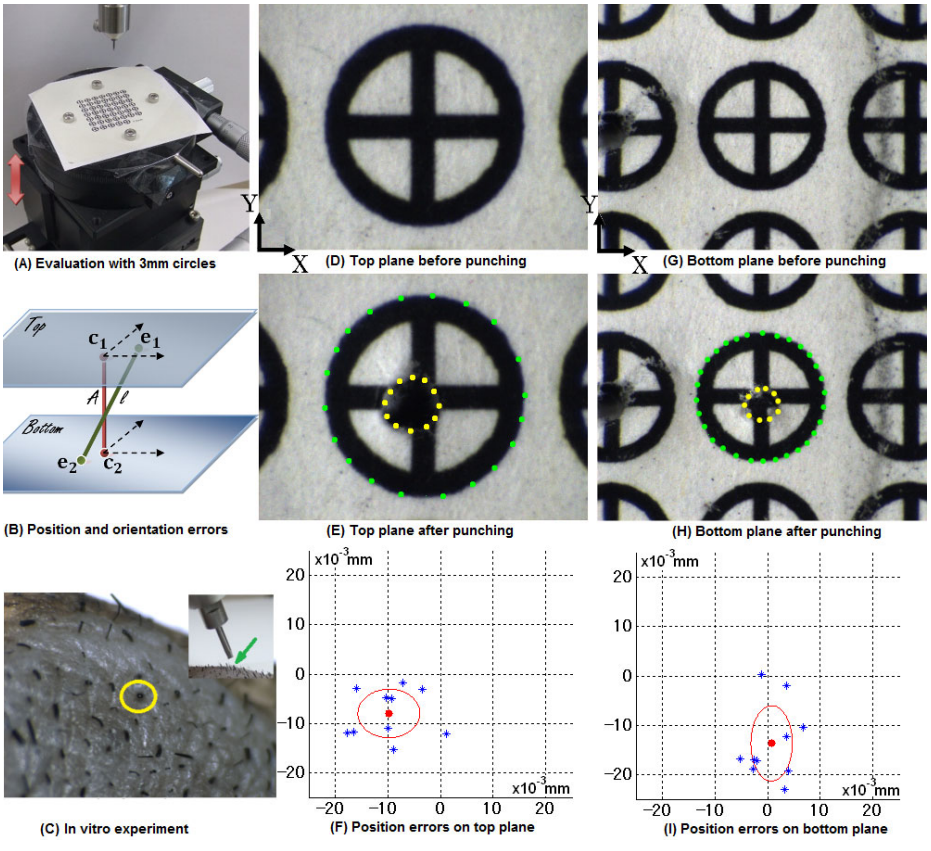


Fig. 4. Evaluation experiments, (A) setup, (B) in vitro experiment, (C) position errors and orientation errors, (D) and (E) images before and after punching on the top plane, (F) position errors on the top plane, (G) and (H) images before and after punching on the bottom plane and (I) position errors on the bottom plane.

to cover the working space from the hair tip above the skin to the papillary bulb tip under the skin, a 6.0 mm thick area was defined for the evaluation. We set the top plane of the working space at the place 3.0 mm higher than the calibration plane and the bottom plane 3.0 mm lower than that plane. Fig. 4D-I show the results on the two planes. Different zoom magnification was applied on each plane (top plane: 38x, bottom plane: 22x), which were estimated from the pixel resolution of the 3.0 mm black circles. The resolution analysis was also used to convert the position errors from pixel to millimeter.

The test was performed 10 times on each plane. A consistent error of -9.8 $\pm$ 5.9 μm in x direction and -8.0 $\pm$ 4.9 μm in y direction on the top plane is observed. A slight large error in y direction which is -13.7 $\pm$ 7.6 μm is found on the bottom plane. The error in x direction on that plane is 0.8 $\pm$ 3.9 μm. The ellipses and their centers in Fig. 4F and 4I give the average errors and the standard deviations in x and y directions.

Orientation error: The orientation error was determined from the average errors (the centers of the ellipse in Fig. 4F and 4I) on the top and the bottom planes. As illustrated in Fig. 4B, if the average error of the top plane is located at e_1 and the error of the bottom plane is at e_2, the angle between the line l of e_1e_2 and the axis A of the planes gives the orientation error. The direction of A can be obtained by connecting two image centers, c_1 and c_2. The orientation error is 0.115 degrees from this calculation.

In vitro experiments were taken on the eyelid part of a 3-month black porcine, which had hair similar to human scalp hair. A hair was first set to the image center (a black dot in Fig. 4C). It was observed that the whole hair could be inserted into the hollow needle with a right angle (small figure in Fig. 4C) after shifting. No hair was actually extracted because of the stiffness of porcine skin.

6 Conclusion and Discussion

In this paper, we proposed a new robotic system for hair harvesting process. A novel end-effector was developed which met the high accuracy requirement in this application. The results demonstrated that the proposed calibration method was efficient and reliable. The system is currently targeting straight hair patients only because of lack of information under the skin. Different imaging modalities, for example high frequency ultrasound or optical coherence tomography, had been tried but failed because of low image qualities. More image modalities will be considered in the future. Mechanical, electrical and software safeties are also needed to be investigated, e.g., force sensors for the needle insertion.

Acknowledgments. This work was partially supported by Grant-in-Aid for Scientific Research (21-09775) of the Japan Society for the Promotion of Science (JSPS). H. Liao was supported by Grant for Translational Systems Biology and Medicine Initiative from Ministry of Education, Culture, Sports, Science & Technology of Japan, and JSPS Grant-in-Aid for Scientific Research (23680049).

References

1. International Society of Hair Restoration Surgery, 2009 RH.Research.: International Society of Hair Restoration Surgery: 2009 Practice Census Results (2009)
2. Bernstein, R., Rassman, W.: Follicular Transplantation. Dermatologic Surgery 23, 771–784 (1997)
3. Rassman, W., Bernstein, et al.: Follicular Unit Extraction: Minimally Invasive Surgery for Hair Transplantation. Dermatologic Surgery 28, 720–728 (2002)
4. Onda, M., Igawa, H., Inoue, K., Tanino, R.: Novel Technique of Follicular Unit Extraction Hair Transplantation with a Powered Punching Device. Dermatologic Surgery 24, 1683–1688 (2008)
5. Gildenberg, P.L.: Hair Transplantation Method and Apparatus. US Patent App. 10/354,836 (2003)
6. Bodduluri, M., Gildenberg, P.L.: Methods for Harvesting Follicular Units Using an Automated System. US Patent App. 11/380,903 (2006)

Direct Surgeon Control of the Computer in the Operating Room

Dumitru Onceanu and A. James Stewart

School of Computing, Queen's University and
Human Mobility Research Centre, Kingston General Hospital

Abstract. This paper describes the design and evaluation of a joystick-like device that allows direct surgeon control of the computer in the operating room. The device contains no electronic parts, is easy to use, is unobtrusive, has no physical connection to the computer, and makes use of an existing surgical tool. The device was tested in comparison to a mouse and to verbal dictation.

1 Introduction

During computer assisted surgery, the surgeon typically delegates control of the computer to an assistant and verbally instructs that assistant to perform operations on the computer. Such verbal communication is slower and can be frustrating to the surgeon if the assistant is not well trained on the computer interface. Verbal communication is also prone to misunderstandings and errors, a problem magnified by verbal ambiguities [1].

This paper describes the design and evaluation of an input device that permits direct surgeon control of the computer. The device consists of a tracked surgical probe (which already exists in the operating room) and a base into which to fit the probe (Figure 1). The surgeon controls the computer by placing the probe into the base and manipulating it like a joystick. The base provides force feedback and permits intuitive clicking with the probe.

A user study was done to evaluate the joystick device in relation to two other input methods: (1) a mouse in a sterile bag and (2) verbal dictation to an assistant. The study considered a set of 1D, 2D, and 3D tasks common to computer interfaces in the operating room.

2 Related Work

It has long been recognized that the computer operator is a weak link in the surgeon-computer interface [2,3]. There have been many attempts to introduce direct surgeon control of the computer into the operating room. Devices for this purpose include static keyboards, foot pedals, touch screens, and sterile bags around conventional devices, such as mice and game controllers.

New technology usually has slow acceptance in the operating room, in part due to human factors limitations [4]. A new device can require a substantial training period for both the surgeon and staff.

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 121–128, 2011.

Fig. 1. The input device is activated when a tracked surgical probe is inserted into the base. Force feedback returns the probe to center and permits clicking.

Foot pedals have wide acceptance and are used as simple switches. But they have severe limitations as they provide no spatial interaction capabilities [5].

A sterile bag wrapped around a mouse or flight controller provides a familiar device to the surgeon, but may be difficult to use as a large, flat area is required, and may increase the risk of infection as the bag might break [6].

A static keyboard is an optically tracked metal plate with markings of various keys on its surface. A keypress is recognized when the surgeon places the tip of a tracked probe onto one of the "keys". These devices are easily sterilized and use tools and tracking that already exist in the operating room, but do not give the surgeon two-dimensional control of the cursor [7].

Some commercial systems use a touch screen with a pen interface. These systems are conceptually easy to use and require no additional equipment. They are, however, placed at some distance from the surgical field and may present a sterility risk from electrostatic transfer of particles from the screen to the surgeon's hand. Touch screens are often used by an assistant to whom the surgeon dictates instructions.

Eye tracking as an input mechanism has been demonstrated for image stabilization and depth recovery in minimally invasive robotic surgery [8]. Characterization of eye movement can also be used to detect stages in a surgical workflow [9].

A promising solution to the surgeon-computer interaction problem comes through a gestural interface in which the surgeon's hands are tracked with computer vision techniques [3,10]. With pointing gestures, the surgeon can move the cursor on screen in a two-dimensional manner. For clicking, the surgeon can either "hold-to-click" by holding their finger motionless for a short period when the cursor is over a button, or "push-to-click" by pushing their finger toward the screen. When the system was evaluated in relation to a mouse, it was found that the mouse could be used to click anywhere on the screen in five seconds, while the gestural interface took 12 seconds. Limitations of the system include hand

fatigue with prolonged use and a lack of definition of the interaction volume, as subjects can stray outside the volume without realizing it.

3 The Joystick Device

The device consists of a base and a surgical probe. The probe is placed into the base to activate interaction and can be removed when needed for other surgical tasks. The base permits the probe to be moved much like a joystick, where the azimuth and elevation angles of the probe determine the direction and speed, respectively, of the cursor on the computer screen. A clicking motion is made by pressing the probe further into the base. A dragging motion is made by pressing the probe into the base, tilting the probe in the direction of motion and returning it to neutral when the motion is finished, then releasing the downward pressure.

Force feedback is provided through a set of magnets in the base and is used to (a) return the probe to a vertical position, (b) provide a tactile "click" when the probe is pressed downward, and (c) return the probe to the "mouse up" position after clicking pressure is released.

The probe, shown in Figure 1, is 30 cm long with a custom-built handle that holds a passive dynamic reference body (Traxtal, Toronto, Canada) at the top, angled 40 degrees from horizontal so as to be visible by the camera in most orientations that it would be used.

The base, shown in Figure 2, is 52 mm in diameter and 80 mm high and contains three main parts: a cylindrical tool insert, a ball, and a socket. An exploded view is shown in Figure 3. The tool insert holds the probe and has a magnet that interacts with magnets in the ball to provide tactile clicking. The ball magnets also interact with magnets in the base to provide force feedback and to return the probe to vertical. The base was manufactured with a Dimension SST 3D printer (Stratasys Inc., Eden Prairie, USA).

The base is fixed to a non-moving surface and is registered by touching the probe to four divots in the upper ring of the socket. The divots define the local coordinate system of the base. If the base is moved, it can be re-registered in ten seconds.

We used the magnets in the bottom of the socket to fix the base to a ferromagnetic surface, but this may not be possible in the operating room and would require that the base be modified to incorporate a clamp. Multiple bases can be used in the same operating room, since the computer can keep track of each base's location and coordinate system.

The device can be sterilized using a low-temperature hydrogen peroxide gas plasma, as is done with the STERRAD system (Advanced Sterilization Products, Irvine, USA). The base can be completely disassembled for sterilization and reassembled in the operating room prior to the operation. Assembly and disassembly take at most a minute each.

During surgery, the location of the tip of the probe is monitored to determine when it is inserted into the base. When this happens, the system changes the state of the probe from a surgical pointing device to a joystick interaction device.

Fig. 2. The base of the joystick device. The probe is inserted into the hole on top of the small cylinder and can be tilted with two degrees of freedom. Pressing the probe down to click causes the small cylinder to move into the ball while still permitting the probe to be tilted.

4 Experimental Evaluation

We evaluated the performance of the joystick device in comparison to two existing surgeon/computer interface methods: a bagged mouse and dictation to an assistant using a touchscreen. The bagged mouse was included as a "baseline" since we expected that it would easily outperform the joystick and dictation methods.

The experimental setup used a Visual Prompting System (VPS) screen on which instructions were presented and a User Interface (UI) screen on which the tasks were performed. A subject would read the task instructions on the VPS screen and would then perform the task (or instruct the assistant to perform the task) on the UI screen, which was two meters away.

The experiment was designed to determine the speed and accuracy of task completion. Three types of tasks were performed: 3D rotations, 2D panning and zooming, and 1D clicking. Three tasks in 3D consisted of rotating a femur to a desired orientation. Three tasks in 2D involved panning X-ray images and moving sliders. Five tasks in 1D included the selection of points on an image and button clicks.

For each task, the subject was shown the task on the VPS screen. They were asked to read and understand the task before signalling that they were ready to perform it. For the mouse and joystick methods, the subject performed the task themselves. For the dictation method, the subject verbally dictated the task to an assistant who could not see the VPS screen. (We recruited members of our laboratory as assistants. An assistant did not participate in more than three evaluations and never participated in two evaluations within two days of each other.)

The test group consisted of seven orthopaedic surgeons (ages 37 to 58) and eleven orthopaedic surgical residents (ages 26 to 36).

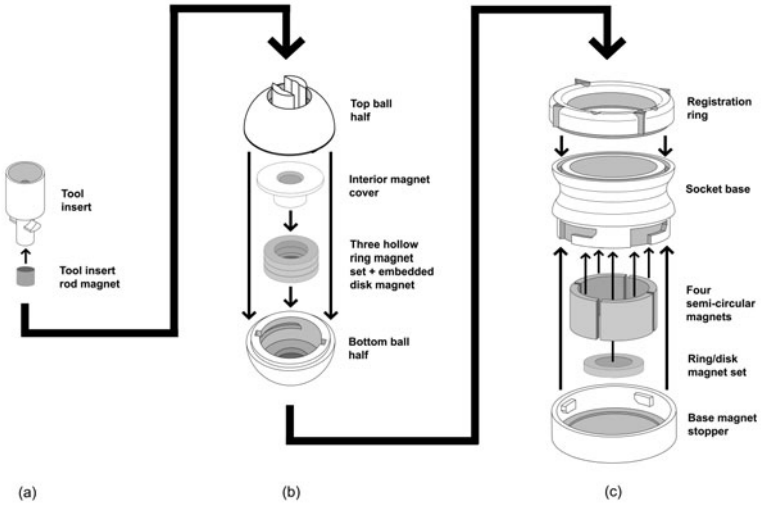


Fig. 3. Exploded view of the base showing (a) the tool insert, (b) the ball, and (c) the socket, along with the internal magnets

Before the evaluation began, the subject was given ten minutes to practice using the joystick. The three methods were tested in a random order. For each method, the same tasks were performed but the order of tasks was randomized. After the evaluation was finished, the user completed a questionnaire. The entire evaluation lasted approximately 30 minutes.

5 Experimental Results

Completion time and accuracy were recorded for all 3D tasks. 3D accuracy was measured as the angular difference between the achieved pose of the femur and the desired pose. The 2D and 1D tasks were always completed successfully, so only completion time is reported for them. We compared performance between mouse and joystick, between mouse and dictation, and between joystick and dictation. A one-sided paired t-test was used to evaluate each hypothesis involving the mouse because we expected the performance of the mouse to be superior on all tasks. A two-sided paired t-test was used for each hypothesis involving the joystick and dictation because we did not know which would be better.

The aggregate results are summarized in Table 1. For the mouse-joystick and mouse-dictation comparisons, the mouse performed significantly faster than the other two methods in all cases. For 3D accuracy, no significant differences were found between the mouse and the other two methods.

For the joystick-dictation comparison times, the joystick was significantly faster for 1D tasks, while dictation was significantly faster for 2D tasks. No significant difference in time was found for 3D tasks, although the trend was in favor of dictation. For 3D accuracy, dictation was significantly better than the joystick (by 1.6 degrees on average).

Table 1. Accuracy and completion times for 1D, 2D, and 3D tasks. The average over all tasks is shown in the "avg" column. The three rightmost columns show the differences between the methods as "average difference ± 95 % confidence interval". A negative difference favors the first method listed in the column heading. P values are shown where significant (p = .00 means p < .005).

	avg	Joystick-Dictation		Mouse-Joystick		Mouse-Dictation	
		avg ± CI	p	avg ± CI	p	avg ± CI	p
1D time (s)	10.5	-3.0 ± 2.6	.02	-5.4 ± 2.0	.00	-8.8 ± 2.2	.00
2D time (s)	15.2	3.9 ± 2.8	.01	-11.7 ± 2.6	.00	-7.8 ± 3.0	.00
3D time (s)	47.9	11.5 ± 13.2		-27.5 ± 10.1	.00	-14.0 ± 10.0	.00
3D accuracy (deg)	9.4	1.6 ± 1.5	.04	0.6 ± 2.7		0.6 ± 2.7	

After completing the 33 tasks, each subject filled out a questionnaire in which they ranked the three methods by "ease of use" (Figure 4 (left)). The mouse was ranked first by all but one subject, who preferred the joystick. The joystick was ranked above dictation by ten of 17 subjects: four of the six surgeons and six of the eleven residents. One surgeon did not complete the questionnaire.

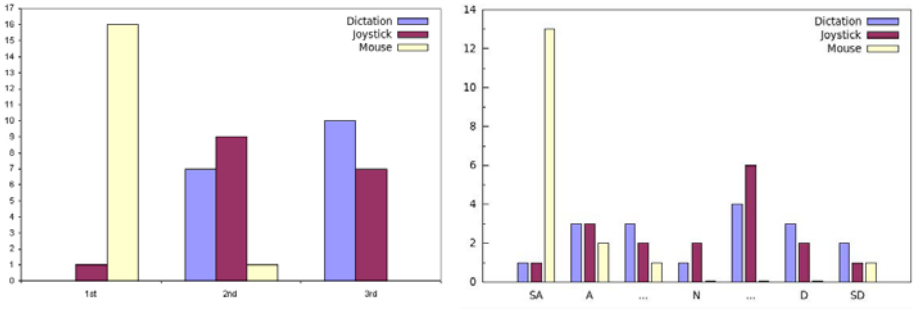


Fig. 4. Left: Rankings of the three methods for ease of use, showing number of respondents on the vertical scale. Right: Responses to the statement "The _____ method was easy to use" on a scale from Strongly Agree (SA) to Strongly Disagree (SD).

The questionnaire also asked subjects to rate each of the methods individually for "ease of use" on a seven-point Likert scale (Figure 4 (right)). The mouse was highly rated by almost all subjects. Overall, the joystick was rated slightly higher than dictation.

A Mann-Whitney test found significance for only mouse-joystick and mouse-dictation at p < .0001 for both ranking and ease-of-use. There was no significance found for joystick-dictation (p = .33 for ranking; p = .81 for ease-of-use).

Following the post-study questionnaire, the experimenter asked the subject for general comments and suggestions. Two surgeons said that they liked the dedicated control aspect of the joystick and related first-hand surgical experiences as to why they preferred to be in full control. Two residents suggested

that the y axis of the joystick should be inverted as is done with handheld game controllers, and two residents suggested that clicking should be done with a trigger button instead of a downward pressing motion. Three subjects mentioned that there was a tendency to overshoot with the joystick, and that they were hesitant to speed up with the joystick in the 3D tasks because they were less familiar with the speed properties of the joystick.

6 Discussion

As expected, the evaluation found that the mouse was faster than both the joystick and dictation methods. The mouse, however, is not readily accepted in the OR because of the need to wrap it in a sterile plastic bag and to find a flat surface in the surgical area upon which to use it. While the mouse was faster, it was not significantly more accurate in 3D tasks.

It was interesting to see, in the progression from 1D to 2D to 3D, that the initial time advantage of the joystick was increasingly lost to dictation (although the 3D times are not statistically significant). Perhaps, as several subjects said, their unfamiliarity with the speed properties of the joystick caused them to move more slowly with the more complex 2D and 3D tasks. This may be due to subjects' lack of familiarity with the joystick's "velocity control" in which cursor *speed* is controlled, rather than cursor *position*, as it is with a mouse or touchscreen. We used speed control as it allows the cursor to be moved an arbitrary distance using the limited range of angles of the joystick. But we now think that the joystick performance might be improved by using position control for certain tasks, such as rotating objects and dragging sliders.

For 3D accuracy, dictation was found to be significantly more accurate than the joystick, by 1.6 degrees on average. For rotating a display to show a particular orientation, 1.6 degrees is probably insignificant when the average error is 9.4 degrees.

The difference in 3D accuracy is likely because the touchscreen, which was used in the dictation method, provided more resolution and responsiveness than the joystick. The joystick's responsiveness could be improved by using an active tracker on the probe, rather than the passive tracker used in the experiments.

In the subjective "ease of use" questions, the mouse was strongly favored, as expected. Surgeons favored the joystick slightly more than did residents, perhaps because the surgeons were more familiar with the problems of dictation in the operating room, and perhaps because the residents had greater gaming experience and had difficulty with the inverted y axis of the joystick.

7 Conclusions

We have described a surgical joystick that gives the surgeon direct control the computer and avoids the difficulties that occur with dictation to an assistant. The device is simple, small, and uses an existing surgical tool to operate. Experimental results show that it is as easy to use as the conventional dictation

method, although somewhat slower for 2D and 3D tasks. Improvements in the joystick software, such as inverting the y axis and using position control for dragging, may overcome this speed difference.

Acknowledgements. Thank you to our subjects for their participation and to the reviewers for their helpful comments. This work is supported by Natural Sciences and Engineering Research Council of Canada (NSERC) grant STPGP 336797 and by the joint NSERC/Canadian Institutes of Health Research CHRP grant 385959.

References

1. Grange, S.: M/ORIS - Medical / Operating Room Interaction System. PhD thesis, Ecole Polytechnique Federale de Lausanne, Lausanne (2007)
2. Visarius, H., Gong, J., Scheer, C., Haralamb, S., Nolte, L.: Man-machine interfaces in computer assisted surgery. Computer Aided Surgery 2(2), 102–107 (1997)
3. Graetzel, C., Fong, T., Grange, S., Baur, C.: A Non-Contact Mouse for Surgeon-Computer Interaction. Technology and Health Care 12(3) (2004)
4. Rivkin, G., Liebergall, M.: Challenges of technology integration and computer-assisted surgery. Journal of Bone and Joint Surgery 91, 13–16 (2009)
5. van Veelen, M.A., Snijders, C.J., van Leeuwen, E., Goossens, R.H.M., Kazemier, G.: Improvement of foot pedals used during surgery based on new ergonomic guidelines. Surgical Endoscopy 17(7), 1086–1091 (2003)
6. Ionescu, A.: A mouse in the O.R. Ambidextrous. Stanford University Journal of Design (2006)
7. Caversaccio, M., Bachler, R., Ladrach, K., Schroth, G., Nolte, L., Hausler, R.: The Bernese frameless optical computer aided surgery system. Computer Aided Surgery 4(6), 328–334 (1999)
8. Mylonas, G., Darzi, A., Yang, G.Z.: Gaze-contingent control for minimally invasive robotic surgery. Computer Aided Surgery 11(5), 256–266 (2006)
9. James, A., Vieira, D., Lo, B., Darzi, A., Yang, G.Z.: Eye-gaze driven surgical workflow segmentation. Medical Image Computing and Computer-Assisted Intervention 10, 110–117 (2007)
10. Grange, S., Fong, T., Baur, C.: M/ORIS - Medical / Operating Room Interaction System. In: International Conference on Multimodal Interfaces, p. 159 (2004)

Learning Gestures for Customizable Human-Computer Interaction in the Operating Room

Loren Arthur Schwarz*, Ali Bigdelou*, and Nassir Navab

Computer Aided Medical Procedures, Technische Universität München, Germany
{schwarz,bigdelou,navab}@cs.tum.edu
http://campar.cs.tum.edu/

Abstract. Interaction with computer-based medical devices in the operating room is often challenging for surgeons due to sterility requirements and the complexity of interventional procedures. Typical solutions, such as delegating the interaction task to an assistant, can be inefficient. We propose a method for gesture-based interaction in the operating room that surgeons can customize to personal requirements and interventional workflow. Given training examples for each desired gesture, our system learns low-dimensional manifold models that enable recognizing gestures and tracking particular poses for fine-grained control. By capturing the surgeon's movements with a few wireless body-worn inertial sensors, we avoid issues of camera-based systems, such as sensitivity to illumination and occlusions. Using a component-based framework implementation, our method can easily be connected to different medical devices. Our experiments show that the approach is able to robustly recognize learned gestures and to distinguish these from other movements.

1 Introduction

Computerized medical systems, such as imaging devices, play a vital role in the operating room (OR). At the same time, surgeons often face challenges when interacting with these systems during surgery. Due to sterility requirements, control terminals are in many cases spatially separated from the main operating site. A typical resulting situation is that a less skilled assistant controls the computer using keyboard and mouse, guided by verbal communication with the surgeon [1,2]. This indirection can be inefficient and cause misundertandings. In addition, surgeons often prefer to have manual control over computerized systems for immediate feedback and, thus, higher precision [1].

We propose a method that allows surgeons to interact with medical systems by means of gestures. Based on the circumstances and the workflow of a particular interventional scenario, a surgeon can define a set of gestures that are most suitable. After demonstrating each gesture to the proposed system, our method learns prior gesture models from the training data (Section 2). These models, termed gesture manifolds, efficiently capture the underlying structure of the movements for each gesture and provide a low-dimensional search space

* Joint corresponding and first authors.

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 129–136, 2011.

for gesture recognition. We model a gesture as a sequence of multiple, smoothly varying body poses, allowing surgeons to adjust continuous parameters. While recognizing gestures, the prior models enable us not only to determine which gesture is performed (categorical control), but also to infer the particular pose within one gesture (spatio-temporal control). Instead of using video cameras to capture gestures, we rely on the data of a few wireless inertial sensors on the surgeon's body. This way, our method can easily handle the challenging conditions of operating room environments, where lighting is highly variable and personnel and equipment cause complex occlusions. Additionally, gestures are recognized regardless of the surgeon's position and orientation. As each inertial sensor can be identified uniquely, gestures can also be assigned to multiple persons, e.g. a surgeon and an assistant, for distributing the interaction workload. By building on a component-based framework implementation, our system enables an easy and dynamic association of learned gestures to the properties of arbitrary intra-operative computer-based systems (Section 3).

We evaluate our gesture-based interaction approach in the scenario of controlling a medical image viewer. Quantitative experiments show that the proposed method can simultaneously recognize up to 18 gestures to a high accuracy from as little as four inertial sensors. The promising results of the usability study encourage a practical application of our method in the operating room.

Related Work. Several authors have recently proposed gesture-based interaction systems for the OR [3,2,4,5]. Graetzel et al. [2] use a stereo camera setup for tracking a surgeon's hand and controlling the mouse pointer on a screen. In [3], hand gestures are recognized from two cameras for controlling a medical image viewer. A similar functionality is presented in [4] using a time-of-flight camera. Guerin et al. [6] use gestures for controlling a surgical robot. Vision-based approaches, such as all above methods, require that gestures are performed in a restricted region seen by a camera. We argue that using wireless inertial sensors for capturing a surgeon's gestures alleviates restrictions of visual systems, e.g. dependence on illumination and line of sight. Inertial sensors have been used for activity [7,8] and gesture recognition [9,10], but the OR has not been addressed. Several authors emphasize the inter-person variability of human gestures, e.g. [7], and propose methods that adapt to person-specific variations in performing a given set of gestures [7]. To provide most flexibility to surgeons, our system allows defining a completely arbitrary set of gestures by only demonstrating one example per gesture. While existing methods typically treat gestures as single commands, such as "click the mouse", the proposed gesture manifold model enables us to automatically recognize the performed gesture and to track the movements within a gesture for fine-tuning continuous parameters. Manifold learning techniques have shown to provide compact, low-dimensional representations of human motion data [11] and have been used for human pose tracking [12,8]. We combine multiple, gesture-specific manifold models and subdivide the embeddings into phases allowing us to assign particular poses to arbitrary parameter settings. Our method also naturally handles the problem of gesture segmentation by means of a predictive confidence measure.

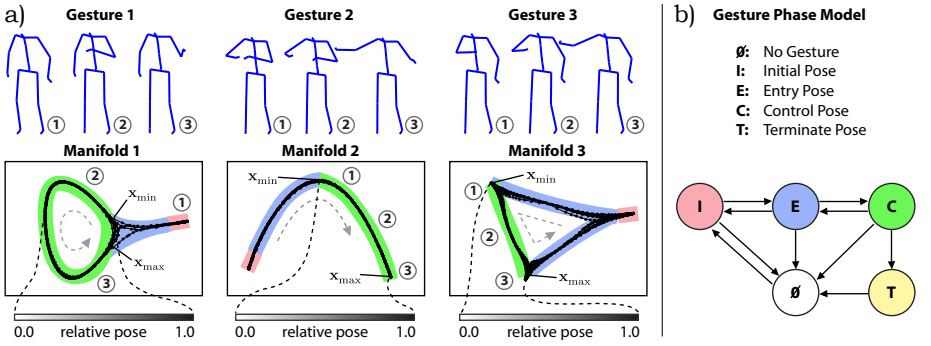


Fig. 1. a) The proposed method recognizes multiple user-defined gestures (*top*) and tracks the relative pose within a gesture by means of learned gesture manifolds (*middle*). The relative gesture pose, given by a value in the interval $[0, 1]$, is used for smooth parameter adjustment (*bottom*). b) Gesture phase model, colors correspond to a).

2 Gesture Recognition Method

The proposed gesture recognition method consists of a training phase, where gesture models are learned from example sensor data for each considered gesture, and a testing phase, where the models are used to recognize gestures from previously unseen sensor data.

Learning Gesture Manifolds. In the training phase, we learn prior models of gestures from sample sensor data. Let N be the number of considered gestures and let $\mathbf{S}^c = [\mathbf{s}_1^c, \ldots, \mathbf{s}_{n_c}^c]$, $1 \leq c \leq N$, be a dataset of n_c labeled sensor measurements. Each vector $\mathbf{s}_i^c \in \mathbb{R}^{d_s}$ consists of four quaternion values per sensor. To obtain a compact parameterization of feasible sensor values, we use Laplacian Eigenmaps, a manifold learning technique [13]. In particular, we map the training data $\mathbf{S}^c$ for each gesture c to a low-dimensional representation $\mathbf{X}^c = [\mathbf{x}_1^c, \ldots, \mathbf{x}_{n_c}^c]$, such that $\mathbf{x}_i^c \in \mathbb{R}^{d_x}$ and $d_x \ll d_s$. Figure 1.a) shows exemplary two-dimensional manifold embeddings for three gestures. The crucial property of the manifold embeddings is that the local spatial distribution of vectors in the original, high-dimensional representation is preserved. In particular, similar sensor measurements will map to close-by embedding points, even if they occur at different times within a gesture. In the testing phase, this makes our gesture recognition method invariant to movement speed.

We relate the space of sensor measurements and the low-dimensional manifold embeddings using kernel regression mappings. The mappings allow projecting new sensor values $\mathbf{s}^*$ to points $\hat{\mathbf{x}}$ in embedding space (*out-of-sample mapping*) and predicting sensor vectors $\hat{\mathbf{s}}$ from given embedding points $\mathbf{x}^*$ (*reconstruction mapping*). Following [8], we define the *out-of-sample mapping* for gesture c as $\hat{\mathbf{x}} = f_c(\mathbf{s}^*) = \frac{1}{\phi^c(\mathbf{s}^*)} \sum_{i=1}^{n_c} k_s^c(\mathbf{s}^*, \mathbf{s}_i^c)\mathbf{x}_i^c$, where $\phi^c(\mathbf{s}^*) = \sum_{j=1}^{n_c} k_s^c(\mathbf{s}^*, \mathbf{s}_j^c)$. We use a Gaussian kernel k_s^c with a width determined from the variance of the training sensor data. The mapping is a weighted average of all manifold embedding points

$\mathbf{x}_i^c \in \mathbf{X}^c$, with the largest weights attributed to points projected from sensor values $\mathbf{s}_i^c$ which are similar to $\mathbf{s}^*$. By interchanging the roles of sensor values and manifold embedding points, we obtain the *reconstruction mapping* $\hat{\mathbf{s}} = g_c(\mathbf{x}^*)$.

Parameterizing Gesture Manifolds. After training, any pose corresponding to one of the learned gestures can be represented as a pair $(c, \mathbf{x})$, where c identifies one of the N manifold embeddings and $\mathbf{x}$ is a point in that embedding. As we allow gestures to have a temporal extent, we also introduce a simple phase model, shown in Figure 1b). Each manifold embedding is subdivided into three phases $\{I, E, C\}$. Phase I indicates the beginning and end of a gesture, phase E contains introductory movements, such as raising a hand, and phase C is actually used for controlling a target system. The points $[\mathbf{x}^c_{\min}, \mathbf{x}^c_{\max}]$ indicating the boundaries of phase C, mapped to a minimal and maximal parameter setting, can be defined by the user in the training phase, e.g. by holding the respective poses for several seconds. The additional phase $\emptyset$ represents poses that do not belong to any of the learned gestures. Phase T is used to terminate parameter adjustment. Our use of the phase model is explained below.

Recognizing and Tracking Gestures. In the testing phase, we employ a particle filter [14] that continuously explores the gesture manifolds to find the manifold index $\hat{c}_t$ and point $\hat{\mathbf{x}}_t$ that best explain the sensor measurements $\mathbf{s}_t$ at any time t. Every particle $\mathbf{p}_t^i = (c_t^i, \mathbf{x}_t^i)$, $1 \leq i \leq n$, represents one pose hypothesis. Initially, all n particles are randomly distributed across the manifold embeddings. In every iteration of the particle filter, we let particles propagate through the manifold embeddings, ensuring that only positions close to the learned embedding points are sampled (see Figure 3). With a certain probability, particles are allowed to switch between different manifold embeddings. We model this probability to be high in embedding space regions that correspond to an idle pose separating gestures. To evaluate the fitness of a particle, we define the observation likelihood $p(\mathbf{s}_t | c_t^i, \mathbf{x}_t^i) \propto \mathcal{N}(g_{c_t^i}(\mathbf{x}_t^i); \mathbf{s}_t, \mathrm{cov}(\mathbf{S}^{c_t^i}))\, \mathcal{N}(f_{c_t^i}(\mathbf{s}_t); \mathbf{x}_t^i, \mathrm{cov}(\mathbf{X}^{c_t^i}))$. The first normal distribution is centered around the observation $\mathbf{s}_t$, giving a high weight to a particle if the sensor value predicted from its position $\mathbf{x}_t^i$ is close to $\mathbf{s}_t$. In the second normal distribution, centered around $\mathbf{x}_t^i$, particles are favored that are close to the projection of $\mathbf{s}_t$ into embedding space. The final estimated gesture index $\hat{c}_t$ is selected as the most frequent index c_t^i among the best-scoring particles. Among these particles, those with $c_t^i = \hat{c}_t$ are used to compute the final position $\hat{\mathbf{x}}_t$ in embedding space.

Temporal Gesture Segmentation. Having estimated $\hat{c}_t$ and $\hat{\mathbf{x}}_t$, we determine the corresponding phase in our gesture phase model (Figure 1b). Phase $\emptyset$, indicating an unknown (or non-gesture) movement, is activated when the prediction confidence $\lambda_t^{\emptyset} = -\log(k_s^{\hat{c}_t}(\mathbf{s}_t, g_{\hat{c}_t}(\hat{\mathbf{x}}_t)))$ falls below a preset threshold. This measure evaluates how closely the estimated state $\hat{\mathbf{x}}_t$, projected to sensor space, matches the true sensor observation $\mathbf{s}_t$. When in phase $\emptyset$, the gesture prediction will likely be incorrect and can be disregarded. Note that our phase model permits transitions into phase $\emptyset$ from any other phase, allowing the user to exit gesture recognition at any time. To identify the *initial* phase I, and thus the beginning of

a gesture, we define $\lambda_t^I = k_s^{\hat{c}_t}(\mathbf{s}_t, \mathbf{s}_0)$, which evaluates the similarity between the sensor measurements $\mathbf{s}_t$ and the idle pose $\mathbf{s}_0$ used for separating gestures. While λ_t^I is above a preset threshold, we assume the idle pose is taken, implying the onset of a gesture. If neither of the phases $\{\emptyset, I\}$ are active, we assume the current phase is one of $\{E, C\}$ and compute a relative pose value $\hat{a}_t \in [0, 1]$ from the manifold position $\hat{\mathbf{x}}_t$ (see Figure 1a). To this end, we transfer the Cartesian coordinate $\hat{\mathbf{x}}_t$ into a polar representation (r_t, θ_t), such that the pole is at the centroid of the embedding points $\mathbf{x}_i^{\hat{c}_t}$, and keep the angular component θ_t. Since the angular representation $[\theta_{\min}^c, \theta_{\max}^c]$ of the points $[\mathbf{x}_{\min}^c, \mathbf{x}_{\max}^c]$ labeled in the training phase is known, we can compute the desired relative pose value as

$$\hat{a}_t = \begin{cases} (\theta_t - \theta_{\min}^{\hat{c}_t})/(\theta_{\max}^{\hat{c}_t} - \theta_{\min}^{\hat{c}_t}) & \text{if} \quad \theta_{\min}^{\hat{c}_t} \leq \theta_t \leq \theta_{\max}^{\hat{c}_t}, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

We define phase C to be active when $\hat{a}_t \neq 0$. By changing the pose within the boundaries of phase C, the user can fine-tune parameters. When a suitable parameter setting has been found, the user can trigger a transition from phase C to the *termination* phase T. The current parameter value is then stored and movements are ignored for a certain amount of time.

3 Component-Based Implementation

In a complex domain such as the OR, it is important to use a proper underlying architecture to achieve a usable gesture-based user interface. In order to practically control parameters of an arbitrary intra-operative device with recognized gestures, we developed a specialized component-based framework. In this model, components encapsulate features of the target systems and expose them to the framework. Such a modular design provides extensibility and therefore a wide range of computer-based devices can be controlled with the proposed gesture-based user interface. Furthermore, to freely customize the behavior of the system for a specific scenario, we implemented the framework according to the data streaming pipeline model. With this model, users can adapt the user interface response to recognized gestures by altering an underlying pipeline graph. This graph contains a set of components as well as the data flow connections (Figure 2). Using visual editing environments and thanks to late binding, the graph can be defined at runtime without further programming.

As shown in Figure 2, we have created separate components for the inertial sensors, the gesture recognition method and a medical image viewer as an exemplary target device. If the gesture recognition component detects the control phase C, a demultiplexer component forwards the relative pose value $\hat{a}_t$ to a separate output signal based on the gesture index $\hat{c}_t$, which further can be bound to any property of the target system. This removes the dependency on other interaction techniques, such as mouse or voice commands, for switching between different properties to control. Additionally, the demultiplexer component blocks further data flow when the current phase is one of $\{\emptyset, I, E, T\}$. This ensures that no modification in the target system will happen when an unknown movement is performed or when a gesture has been terminated.

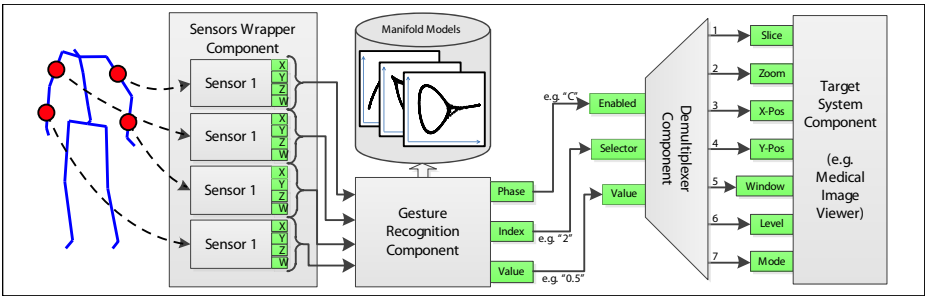


Fig. 2. Component graph for the proposed system. User movements, captured by the inertial sensors, are forwarded through the gesture recognition component, relying on learned manifold models, to properties (green) of the medical target system.

4 Experiments and Results

To evaluate our gesture-based interaction method for the OR, we conducted qualitative and quantitative experiments. We used two to six Colibri Wireless orientation sensors (Trivisio GmbH) attached to the arms of our testing persons. Gestures we defined for evaluation included moving an arm horizontally or vertically, tracing out circles, or other movements with one or both arms. Our implementation with Matlab and C++ components runs in realtime.

User Study. Tests with 10 subjects were performed to assess the usability of the proposed method with the medical image viewer as an exemplary target system. We asked each person to localize a stent bifurcation within a volumetric CT dataset using 6 personalized gestures (Figure 3). The gestures were assigned to the main parameters of the image viewer, such as scaling, contrast, slice number, etc. Average user answers in a questionnaire consisting of five questions are given in Figure 4 a). Very positive feedback was given to the wearability of the sensors, the responsiveness and the achieved precision of the system.

Gesture Recognition. The gesture recognition accuracy of our method was evaluated systematically on a dataset of 18 different gestures, each recorded four times. We created labeled sequences of multiple gestures in a row, each between 1000 and 5000 frames. Experiments were performed in a cross-validation manner, each time using one of the sequences for training and one of the others for testing. While measuring the percentage of frames with correctly recognized gestures, we varied the number of simultaneously trained gestures and the number of inertial sensors. As shown in Figure 4 b), best gesture recognition rates were achieved with six inertial sensors. In this case, 95% of all frames were correctly recognized with a set of four gestures, and 88% when the method was trained on 18 gestures. Although the recognition rates decrease for less than six sensors, even only two sensors (one on each arm) yield correct recognition rates above 80% for all considered numbers of gestures. Our method thus scales well with respect to the number of gestures to be recognized simultaneously. However, we expect that less than 10 gestures need to be distinguishable in practical applications.

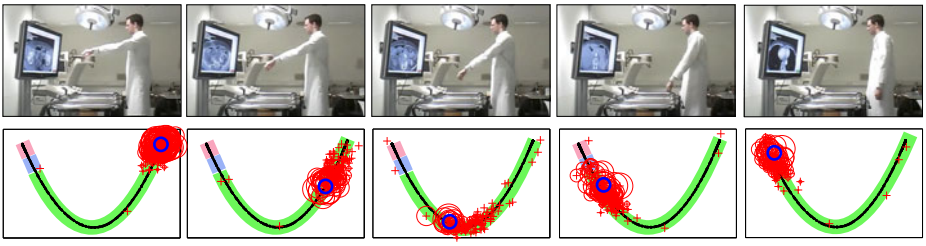


Fig. 3. *Top row*: Sample images of a test subject performing one of the learned gestures. *Bottom row*: Manifold embedding for the same gesture with distribution of particles, shown in red, for each of the above images. The point $\hat{\mathbf{x}}_t$ is given by a dark circle.

Gesture Segmentation. It is crucial for a gesture-based interaction method to distinguish instances of learned gestures from arbitrary other movements. Using six inertial sensors, we varied the threshold associated with the confidence measure $\lambda_t^{\emptyset}$ and measured the percentage of frames with gesture movements recognized as such (true positive rate, TPR) and the percentage of non-gesture frames wrongly identified as gestures (false positive rate, FPR). We trained the method on different numbers of gestures and randomly created sequences where only one of multiple gestures was known to the method. Figure 4 c) shows the resulting ROC curves. The best combination of high TPR and low FPR was achieved in the setting with four gestures. In this case, above 90% of gestures were detected with less than 10% of false positives. Although distinguishing non-gesture movements becomes more difficult when many gestures are learned simultaneously, detection results remain reasonable, even for 18 learned gestures.

5 Discussion and Conclusion

In this paper, we proposed a novel gesture-based interaction method for the OR using wireless inertial sensors. During a training phase, manifold models

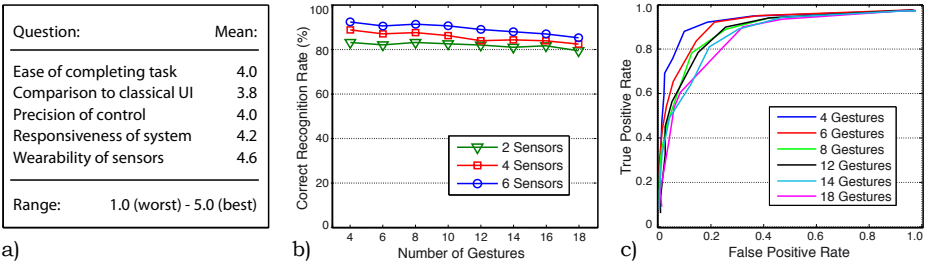


Fig. 4. a) Average questionnaire results after qualitative user study with 10 test subjects. b) Correct gesture classification rates for various settings. c) ROC curves for automatic differentiation of learned gestures from non-gesture movements.

are learned for each considered gesture based on the sensor data. This allows customizing gestures for each individual user and for different workflow phases, considering different constraints. Our evaluation shows promising results of using this method in our experimental setup. While the automatic differentiation between learned gestures and other movements achieves true positive rates above 90%, an additional tool, e.g. a pedal or voice command, could be used to activate and deactivate automatic gesture-based control within critical workflow stages. We will concentrate on these options in our future work.

References

1. Johnson, R., O'Hara, K., Sellen, A., Cousins, C., Criminisi, A.: Exploring the potential for touchless interaction in image-guided interventional radiology. In: ACM Conference on Human Factors in Computing Systems, pp. 1–10 (January 2011)
2. Graetzel, C., Fong, T., Grange, S., Baur, C.: A non-contact mouse for surgeon-computer interaction. Technology and Health Care 12(3), 245–257 (2004)
3. Kipshagen, T., Graw, M., Tronnier, V., Bonsanto, M., Hofmann, U.: Touch-and marker-free interaction with medical software. In: World Congress on Medical Physics and Biomedical Engineering 2009, pp. 75–78 (2009)
4. Soutschek, S., Penne, J., Hornegger, J., Kornhuber, J.: 3-d gesture-based scene navigation in medical imaging applications using time-of-flight cameras. In: Computer Vision and Pattern Recognition Workshops (April 2008)
5. Wachs, J.P., Stern, H., Edan, Y., Gillam, M., Feied, C., Smith, M., Handler, J.: A real-time hand gesture interface for medical visualization applications. Applications of Soft Computing, 153–162 (2006)
6. Guerin, K., Vagvolgyi, B., Deguet, A., Chen, C., Yuh, D., Kumar, R.: ReachIN: A modular vision based interface for teleoperation. In: SACAI Workshop (2010)
7. Liu, J., Zhong, L., Wickramasuriya, J., Vasudevan, V.: uWave: Accelerometer-based personalized gesture recognition and its applications. Pervasive and Mobile Computing 5(6), 657–675 (2009)
8. Schwarz, L.A., Mateus, D., Navab, N.: Multiple-activity human body tracking in unconstrained environments. In: Perales, F.J., Fisher, R.B. (eds.) AMDO 2010. LNCS, vol. 6169, pp. 192–202. Springer, Heidelberg (2010)
9. Hartmann, B., Link, N.: Gesture recognition with inertial sensors and optimized DTW prototypes. In: IEEE Conference on Systems Man and Cybernetics (2010)
10. Kela, J., Korpipää, P., Mäntyjärvi, J., Kallio, S., Savino, G., Jozzo, L., Marca, S.: Accelerometer-based gesture control for a design environment. Pers Ubiquit Comput. 10(5), 285–299 (2006)
11. Elgammal, A., Lee, C.S.: The role of manifold learning in human motion analysis. In: Rosenhahn, B., Klette, R., Metaxas, D. (eds.) Human Motion. Computational Imaging and Vision, vol. 36, pp. 25–56. Springer, Netherlands (2008)
12. Jaeggli, T., Koller-Meier, E., Gool, L.V.: Learning generative models for multi-activity body pose estimation. International Journal of Computer Vision 83(2), 121–134 (2009)
13. Belkin, M., Niyogi, P.: Laplacian eigenmaps for dimensionality reduction and data representation. Neural Computation 15(6), 1373–1396 (2003)
14. Isard, M., Blake, A.: Condensation—conditional density propagation for visual tracking. International Journal of Computer Vision 29(1), 5–28 (1998)

Interventional Radiology Robot for CT and MRI Guided Percutaneous Interventions

Nikolai Hungr[1], Céline Fouard[1], Adeline Robert[1],
Ivan Bricault[2], and Philippe Cinquin[1]

[1] UJF-Grenoble 1 / CNRS / TIMC-IMAG UMR 5525 (GMCAO), Grenoble, F-38041, France
{nikolai.hungr,celine.fouard}@imag.fr
[2] Service Central de Radiologie et Imagerie Médicale, CHU Grenoble,
38700 La Tronche, France

Abstract. This paper introduces a new patient-mounted CT and MRI guided interventional radiology robot for percutaneous needle interventions. The 5 DOF robot uses ultrasonic motors and pneumatics to position the needle and then insert it progressively. The needle position and inclination can be registered in the images using two strategically placed fiducials visible in both imaging modalities. A first prototype is presented and described in terms of its sterilization, CT and MRI compatibility, and precision. Tests showed that 1) it is entirely sterilizable with hydrogen peroxide gas, 2) no image artifacts or deformations are noticeable in the CT and MRI images, 3) does not affect the SNR of MR images, and 4) its mechanical error is less than 5mm.

Keywords: Interventional radiology, Robotics, CT, MRI, Image guided, Ultrasonic motors, Magnetic resonance compatibility, Sterilization.

1 Introduction

Percutaneous interventional radiology (IR) is a field that consists of diagnostic and therapeutic procedures such as image-guided biopsies, cancer treatment by radiofrequency ablation, cryoablation and high-intensity focused ultrasound (HIFU), abscess and pleural drainage, and vertebroplasty. In 2008, 11.5 million IR acts were realized in the US, of which 35% where non-cardiovascular. The popularity of these procedures has incited a growing interest in robotic-assisted techniques.

Of the primary imaging techniques used in IR, CT and MRI are of particular interest due to their ability to view any part of the body in three-dimensions and with typically higher definition. Both modalities have their advantages. CT images have high resolution whereas MRI is non-irradiating, has better contrast density for soft tissues, and allows for thermal imaging for cryoablation for example, as well as non-invasive real time imaging that introduces the possibility of motion tracking and compensation. These differences make CT and MRI imaging complementary, making IR interesting in both modalities.

CT-guidance is a well established technique for percutaneous needle interventions. Some of the main difficulties encountered in conventional CT-guided IR are accessibility due to constrained tunnel dimensions, lengthy and iterative steps that

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 137–144, 2011.

affect patient comfort, difficulty for the clinician to visualize the needle trajectory on the patient, as image planning is done in the distant control room, and important complications when inclining the needle in two directions. MRI-guidance, on the other hand, remains uncommon in clinical practice due to cost, tunnel size and material compatibility constraints.

In this paper, we present a new, patient-mounted multimodal, MRI and CT compatible interventional radiology robot based on our previously published LPR robot [1]. The primary goal of the robot is to increase the scope of percutaneous interventions available to clinicians by 1) increasing needle insertion accuracy by allowing accurate location of needle position and inclination in the images combined with progressive needle insertion, 2) allowing more complex 3DOF approach angles (between the ribs or deep insertions, for example), 3) reducing the number of image acquisitions required (and hence decreasing radiation doses in CT), 4) improving intervention efficiency, and 5) offering the possibility of MR-guided interventions (for example for patients with certain contraindications to CT contrast agents, or for better visualization of equal-density tumors, etc.).

2 State of the Art

Most of the difficulties described in the previous paragraph could be addressed through robotic assistance, and indeed, a multitude of systems have been described in the literature. These systems can be classified into two categories: bed-mounted robots and patient-mounted robots. An example of the first class of robots is the AcuBot [2][3], a CT guided IR robot, which uses a bridge-like macro-micro architecture to insert a needle with 1-2mm precision. Another similarly bridge-mounted robot is the MRI-guided Innomotion system, which uses pneumatic actuators coupled with optical sensors to drive the robot, while needle insertion remains manual, with sub-millimetric accuracy [4]. A third example is the B-RobII system which uses two parallel fingers to align the needle above a target, allowing the clinician to insert the needle manually, with a precision of 0.88mm [5].

Patient-mounted systems have the difference of being smaller and able to move with the patient. The effect of motion due to patient discomfort as well as the motion of the insertion point on the patient's skin during breathing can thus be minimized to a certain extent. The CT-Bot is one such robot [6] that uses ultrasonic motors to power a parallel-structure with 5 degrees of freedom (DOF). It uses a single line-fiducial mounted on the base of the robot for registration with the CT images, allowing an accuracy of <5mm at typical needle insertion depths. The Robopsy is another example of such a CT-guided robot that has 3 DOF for needle orientation and insertion [7], while positioning at the skin insertion point is manual. Neither of these robots are MRI compatible.

The systems described above show the important potential benefit that robotic-assistance can have for improving the quality and feasibility of percutaneous IR procedures. The Light Puncture Robot (LPR) [1] developed in our laboratory, consisting of a patient-mounted needle insertion module stretched between four straps can be pulled individually to control the needle position. The major contribution of this project compared to other existing systems is its compatibility with both CT and MRI guided percutaneous procedures, significantly enlarging its clinical scope.

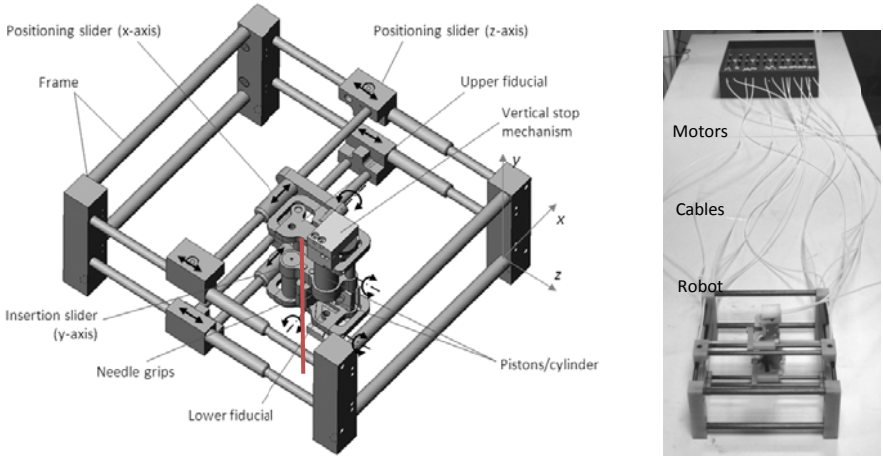


Fig. 1. New robot architecture. The needle is denoted by a vertical red line in the left image.

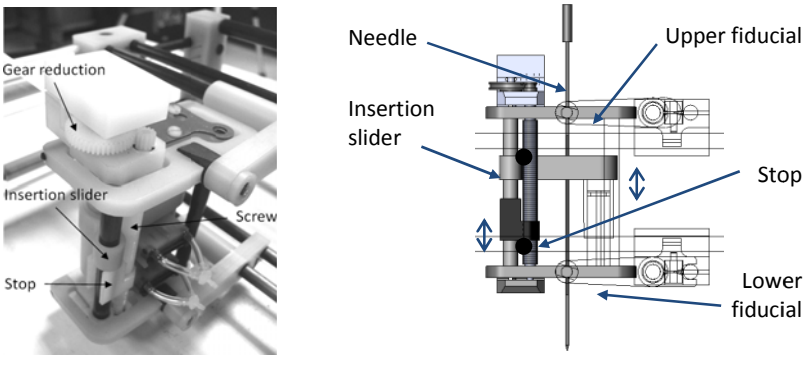


Fig. 2. Stop mechanism for setting the insertion stroke

Although the animal testing reported in [1] showed the feasibility and accuracy of the LPR in a CT-guided environment, a few limitations were revealed to the design. Firstly, although the needle insertion module was light and relatively compact, the size of the surrounding frame limited the size of patient that could fit inside. Additionally, the pneumatic actuators resulted in slow motions and significant noise during pressure-release.

This paper introduces the new LPR design, along with some preliminary precision and compatibility results, and sterilization tests as we prepare for phantom and mock patient trials.

3 Robot Description

The robot architecture is shown in Figure 1. It is based on a 5 DOF parallelogram design, in which the needle is held by two parallel platforms, 85 mm apart, that enable

a maximum planar translation of 135mm and a maximum needle inclination of ±60° in one direction and ±15° in the other. The needle is held by two independent grippers the upper one being mounted on a vertical slider and providing the needle insertion force, as shown in Figure 2. The frame is strapped to the patient's body to follow external movements.

Needle insertion is done incrementally by sequentially activating the two needle grippers and the insertion slider (Figure 2). The stroke distance (adjustable between 0 and 40mm) is regulated by a controllable stop that blocks the vertical slider's path. The stop height is regulated by a screw connected to a set of reduction gears.

Translation of the four x and z sliders is achieved through four Shinsei USR60-E3N ultrasonic motors. Their torque (max. 1Nm) and speed (rated 100rpm) ratings are easily sufficient for the low friction and low reduction characteristics of the robot. As studies have shown that these motors are capable of affecting an MRI image when powered inside the tunnel [8], it was decided to house them in a separate container placed on the scanner or MRI bed, at the patient's feet (see Figure 1, right). They are connected to the robot by 1.5m long cables and housing, ensuring that they are never inside the tunnel. Each motor activates a pair of rack and pinions, which in turn pull the cables to translate the respective robot slider. The cable housing is made of Teflon, to reduce cable friction, while the cables are made of low stretch 0.4mm diameter Spectra thread, typically used in archery. Separating the motors from the robot also reduces the robot's weight on the patient's body.

While the insertion stop mechanism is controlled by an ultrasonic motor, the insertion and retraction of the insertion slider as well as the two needle grippers are powered by pneumatics. This was chosen in order to give sufficient power and speed to these vital elements of the robot. At a typical hospital air pressure of 4 bars, the needle insertion force is >50N, while the gripping force is 8N. The latter increases to 13N at 5 bars, which is largely enough for typical needle insertion forces. Air pressure can be provided either by the hospital air supply that is usually found in or near the imaging rooms, by bottled medical air, or by a small electrical pump, the latter two being, of course, located outside the imaging room, particularly in the case of MRI.

All the robot and motor unit materials consist of a combination of CT and MRI-compatible delrin, epoxy resin, carbon fiber and nylon materials. Fiducials, mounted on both the parallel platforms and both visible in CT and MRI, allow for direct tracking of the needle position and inclination, instead of just the robot frame, as is the case in other robots. This increases accuracy and reliability.

The robot control box, located outside the imaging room, houses the motor controllers and pneumatic solenoid valves. The electronic cables and air hoses pass through the service hatch of the scanner/MRI room. The PC used to control the robot is connected to the hospital network, allowing it to obtain the images coming from the imager.

4 Robot Evaluation

4.1 Sterilization

Along with the functional aspects of the design, sterilization is a vital element to consider. Although not yet fully incorporated into this first prototype, a study of its

sterilizability has been carried out. Percutaneous IR needle interventions can be considered low risk procedures, as they occur in the open environment of the scanner and MRI rooms, and as the inherent risk of a single needle insertion is relatively small. The robot's primary condition for sterilization must be that there be no risk of contact between a non-sterile component of the robot and the needle or the clinician.

Since the needle is located within the confines of the robot's frame, requiring the clinician to reach inside to handle the needle, and since there are many moving elements that would be difficult to simply cover with sterile covers or drapes, it was decided to sterilize the entire robot and part of the cables leading to the motor unit. Any potential air leaks coming from the pneumatic system were deemed irrelevant, as the air in the imaging room is untreated, ambient air in any case.

Due to the numerous moving parts, with 14 cables running in and out of their 1.5mm inner diameter cable housings, and numerous heat sensible materials, sterilization by hydrogen peroxide gas (such as the STERRAD® system, from Advanced Sterilization Products) was chosen. Three aspects for successful sterilization were explored.

First, to fit into the smallest typical sterilization chamber (170 x 300 x 600mm) it is necessary to disconnect the robot from the motor unit. Although not yet incorporated into the prototype described here, a cable connector has been design. This will allow the entire robot and nearby cables to be placed in the sterilization chamber at once.

Second, as the compatibility of carbon fiber with hydrogen peroxide gas was unkown, we decided to test this compatibility. We passed two carbon fiber samples through 25 normal sterilization cycles using the STERRAD® NX machine available to us in our partner hospital (cycle time = 38 min; temperature = 45˚C; pressure = 1.013 bar). Visual microscopic comparison of the surfaces of these samples with non-sterilized samples, showed no evidence of surface decomposition, microfractures, or delamination of the fibers. Bend to fracture tests on these sample bars also showed no noticeable change in fracture strength and fracture mode. As the robot frame undergoes minimal loads during operation, these results were deemed satisfactory.

Third, as temperature is not an acting sterilization agent in the sterilization technique chosen, it is necessary that all parts of the robot be reachable by the gas. This is especially the case for the cables hidden inside the cable housing, which during robot operation will emerge within close range of the needle, as they are pulled in and out. To verify the successful sterilization inside the cable housing, an experiment was undertaken in which 4 Spectra cables were impregnated with Geobacillus Stearothermophilus spores (standard bacterium used for validation of sterilization studies, including autoclave and gas-based techniques) and inserted into 30cm lengths of cable housing. Three of these were passed through a single STERRAD® sterilization cycle, along with a batch of control spores in an open Eppendorf tube. The three sterilized samples, the control spores and the one non-sterilized sample, were then put into cultivation at 60˚C, and the optical densities (OD) were measured using a spectrometer for each cultivation solution at 24h and 72h. At both instances, only the non-sterilized sample showed any sign of bacterial proliferation (OD = 0.952, compared to OD < 0.014 for sterilized samples), confirming that the cables were successfully sterilized. As the cable housing will be disconnected from the robot during sterilization, this experiment was deemed appropriate and conclusive.

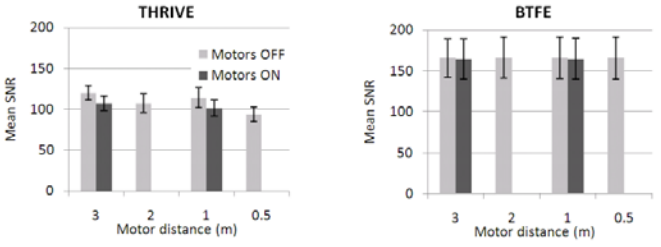


Fig. 3. Average SNR for each set of THRIVE and BTFE images taken during MRI compatibility tests. Motor distance was measured from the opening of the MRI tunnel.

4.2 CT and MRI Compatibility

To verify the CT, and in particular MRI compatibility of our prototype, we conducted three experiments. In the first test, we imaged the robot in a Siemens Somatom Sensation 16 CT scanner at a resolution of 512x512 and a slice thickness of 1mm. The images showed no signs of streaking or beam hardening, artifacts that could affect the clinical image quality. Also, compared to the previous LPR design, this robot stands on four legs, bringing the bulk of the robot away from the patient's skin, making it easily distinguishable in the CT images.

In the second test, we measured the signal to noise ratio (SNR) in a series of MRI image volumes, with the robot in place and with the motor unit at increasing proximity to the MRI tunnel (3m, 2m, 1m, and 0.5m). The SNR was measured in exactly the same manner as described in [1] and [8], using a phantom bottle in a Philips Achieva 3.0T MRI scanner. BTFE[1] and THRIVE[2] sequences, typically used during IR, were taken. The results of the various measurements made are shown in Figure 3 and are within the ranges reported in [1]. We note that all signal to noise ratios are comparable and acceptable.

In the third test, using the same images as in the second test, the upper and lower robot fiducials were segmented using identical thresholds for all the image volumes. Two measurements were made and are summarized in Table 1: the average variation (expressed as a standard deviation) of the center of motion (COM) for each segmented image and the average volume of segmented pixels for all segmented fiducials. The variations of both measurements did not show any sort of pattern based on the motor distance or state. Note that in this test, the registration of the segmented image COM to the true physical robot COM was not considered, as our primary interest here was the effect of the robot and motors on image quality.

Although the two MRI tests presented are preliminary and will have to be furthered on actual patients, it is important to note that 1) the motors will never be actuated during image acquisition, and 2) the motors will never be inside the imaging tunnel. The results were therefore very encouraging.

[1] Balanced Turbo Field Echo: Slice thickness = 5mm, resolution = 256x256, slices = 15, slice spacing = 6.5mm, TR = 2.62ms, TE = 1.31ms, flip angle = 45˚.

[2] T1 Weighted High Resolution Isotropic Volume Exam: Slice thickness = 4mm, resolution = 384x384, slices = 80, slice spacing = 2mm, TR = 3.03ms, TE = 1.42ms, flip angle = 10˚.

Table 1. Average variations (expressed as standard deviation) of the segmented upper and lower fiducials for motor distances of 0.5m (motors OFF), 1m (OFF and ON), 2m (OFF) and 3m (OFF and ON) from the MRI tunnel opening. Note: ON = in motion (i.e. worst case).

	Average variation of pixel volume		Average variation of COM (mm)	
Fiducial:	Upper	Lower	Upper	Lower
BTFE	8%	7%	0.75	0.30
THRIVE	15%	15%	1.02	0.47

4.3 Robot Accuracy

A preliminary test has been run on the robot to determine its needle positioning error before advancing to more in depth needle insertion experiments on phantoms and mock patient trials. The test involved moving the robot inside an MRI imager and comparing images taken before and after the motion. By segmenting the fiducials in both volumes, the distance moved by the robot can be compared to the commanded distance. This comparison gave us an idea of the total error, including robot motion and imaging errors. Segmentation of the images was done manually, using smoothing, thresholding and basic binary processing to isolate the fiducials and then to calculate their centers of mass (COM) within each image. The latter was used to calculate the distance moved. From a total of six repeated 50mm x and y translations, the resulting error was found to be 4mm (SDev = 1.1) for the upper fiducial and 5mm (SDev = 1.2) for the lower fiducial.

These errors come from two sources. First, the images used in this analysis were of fairly low resolution (384 x 384 pixels, 100 slices, 0.9766mm pixel size, 2mm image spacing), significantly affecting the reported error. Second, the mechanical stretching of the cables during motion, result in some hysteresis between the motor movement and the respective slider movement 1.5m away. This effect was minimized through material selection (use of stiff, low friction materials), but is impossible to eradicate completely. To decrease this error, we are looking, on one hand, into a better way of equalizing the cable tension throughout the robot, and on the other to develop an algorithm to predict the hysteresis.

5 Conclusion

We are continuing development on this prototype in order to carry out phantom needle insertion tests in both CT and MRI machines, as well as further mock patient tests in the MRI machine. Specific research aspects include the development of automatic fiducial segmentation algorithms, their incorporation into the robot kinematics, and ergonomic, size and actuation redundancy improvements (in particular, to increase the safety of the needle insertion module and to ensure no unwanted motion of the linear slides during needle insertion), robot-to-patient fixation design, and hysteresis tests on the cables, to prepare the robot for clinical trials.

Acknowledgments. We would like to acknowledge Nabil Zemiti for spearheading the original design of this robot, Jean Pierre Alcaraz for his expertise with the bacterial study and the departments of sterilization, security and hygiene of the Grenoble University Hospital.

References

1. Zemiti, N., Bricault, I., Fouard, C., Sanchez, B., Cinquin, P.: LPR: A CT and MR-Compatible Puncture Robot to Enhance Accuracy and Safety of Image-Guided Interventions. IEEE/ASME Transactions on Mechatronics 13(3), 306–315 (2008)
2. Stoianovici, D., Cleary, K., Patriciu, A., Mazilu, D., Stanimir, A., Craciunoiu, N., Watson, V., Kavoussi, L.: AcuBot: a robot for radiological interventions. IEEE Transactions on Robotics and Automation 19(5), 927–930 (2003)
3. Cleary, K., Watson, V., Lindisch, D., Taylor, R.H., Fichtinger, G., Xu, S., White, C.S., Donlon, J., Taylor, M., Patriciu, A., Mazilu, D., Stoianovici, D.: Precision placement of instruments for minimally invasive procedures using a “needle driver” robot. Int. J. Med. Robot. 1(2), 40–47 (2005)
4. Melzer, A., Gutmann, B., Remmele, T., Wolf, R., Lukoscheck, A., Bock, M., Bardenheuer, H., Fischer, H.: INNOMOTION for percutaneous image-guided interventions: principles and evaluation of this MR- and CT-compatible robotic system. IEEE EMB Magazine 27(3), 66–73 (2008)
5. Kronreif, G., Fürst, M., Ptacek, W., Kornfeld, M., Kettenbach, J.: Robotic System for Image Guided Therapie - B-RobII. In: RAAD Workshop, BFD-022 (2006)
6. Maurin, B., Bayle, B., Piccin, O., Gangloff, J., de Mathelin, M., Doignon, C., Zanne, P., Gangi, A.: A Patient-Mounted Robotic Platform for CT-Scan Guided Procedures. IEEE Transactions on Biomedical Engineering 55(10), 2417–2425 (2008)
7. Walsh, C., Hanumara, N.C., Slocum, A.H., Shepard, J.A., Gupta, R.: A Patient-Mounted, Telerobotic Tool for CT-Guided Percutaneous Interventions. J. of Med. Devices 2(1), 011007.1–011007.10 (2008)
8. Fischer, G.S., Krieger, A., Iordachita, I.I., Csoma, C., Whitcomb, L.L., Fichtinger, G.: MRI Compatibility of Robot Actuation Techniques – A Comparative Study. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 509–517. Springer, Heidelberg (2008)

Spatio-Temporal Registration of Multiple Trajectories

Nicolas Padoy and Gregory D. Hager

The Johns Hopkins University, Baltimore, MD, USA
{padoy,hager}@jhu.edu

Abstract. A growing number of medical datasets now contain both a spatial and a temporal dimension. Trajectories, from tools or body features, are thus becoming increasingly important for their analysis. In this paper, we are interested in recovering the spatial and temporal differences between trajectories coming from different datasets. In particular, we address the case of surgical gestures, where trajectories contain both spatial transformations and speed differences in the execution. We first define the spatio-temporal registration problem between multiple trajectories. We then propose an optimization method to jointly recover both the rigid spatial motions and the non-linear time warpings. The optimization generates also a generic trajectory template, in which spatial and temporal differences have been factored out. This approach can be potentially used to register and compare gestures side-by-side for training sessions, to build gesture trajectory models for automation by a robot, or to register the trajectories of natural or artificial markers which follow similar motions. We demonstrate its usefulness with synthetic and real experiments. In particular, we register and analyze complex surgical gestures performed by tele-manipulation using the da Vinci robot.

1 Introduction

Time series data is gaining importance in medicine and poses new challenges. Whereas historically the spatial aspect of data was analyzed, for instance to build atlases, we now need to consider both the spatial and the temporal aspects together. Trajectories are natural spatio-temporal data descriptors. Our interest in this paper focuses specifically on tool trajectories that allow to model and analyze the performance of surgeons. Models of surgical performance are used to develop training systems [1] and evaluation methods [2]. With the rapid advancement of minimally invasive surgery, there is a growing need for such systems that can improve the learning curve of trainees. Surgical models can also be used to segment gestures [3] and to recognize the surgical workflow [4], in order to develop context aware assistance systems. More recently, expert gesture information has been used in robotics to make surgical robots learn how to perform complex movements like knot tying [5].

Gesture datasets generally consist of tool trajectory data recorded from different users, possibly operating in different environments. As a result, the data contains variations. For example, rigid spatial transformations arise from different patient and system locations. Different operator styles and environment constraints cause not only rigid and deformable spatial deformations, but also non-linear speed differences in the executions. A common way to recover spatial differences is to use markers. However, such a setup may be cumbersome in some scenarios or non-available in some datasets.

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 145–152, 2011.

Analyzing multiple trajectories representing the same gesture requires therefore the recovery of both the rigid spatial transformations and the non linear time warpings. This way, multiple instances can be both spatially and temporally superimposed, allowing the comparison of variations solely caused by style, experience and environment specificities. Note that this recovery also provides the temporal synchronization between videos observing the scenes. Such information is very valuable for surgeons, as it provides an intelligent way to mine surgical databases, e.g. to extract and compare side by side surgical performances during training seminars.

In this paper, we introduce two cost functions, one stating the spatio-temporal registration problem for a pair of trajectories, and the second stating the spatio-temporal registration problem for multiple trajectories. We then propose optimization approaches that combine dynamic time warping (DTW) [6] with rigid spatial registration [7]. In the case of multiple trajectories, the registration approach assumes that they are all transformations of a generic gesture template to be computed. It iteratively computes the rigid transformation and temporal warping between each trajectory and the template, which is updated. The recovery of the spatial transformations permits the computation of an accurate template that keeps the same shape as all trajectories. Also, rotation invariant features like curvature are then not necessary to drive the temporal synchronization.

Usually, a surgical gesture spatially overlaps over time within a small 3-dimensional volume. Therefore, methods like iterative closest point [8], which uses only spatial information, or dynamic time warping [6], which does not model spatial transformations, are not accurate for registration. We show that the spatio-temporal registration provides better results. Our approach share similarities with canonical time warping (CTW) [9], which uses DTW and canonical correlation analysis to jointly synchronize and extract common features from a pair of time-series having possibly different dimensions. Two main differences are: 1) we address the joint registration of multiple time-series and 2) due to the nature of our problem where time-series represent trajectories, we estimate a rigid transformation between two time-series instead of projection matrices for each of them. A synthetic experiment in section 3 will show how the choice of the appropriate modeling for the transformation affects the registration.

The remainder of this paper is as follows: we present our approach in section 2 and show synthetic and real experiments in section 3. Conclusions are given in section 4.

2 Methods

Sections 2.1 and 2.2 present two standard approaches to register two trajectories either temporally or spatially. In section 2.3, we state the spatio-temporal registration problem for two time-series and explain how to combine the two previous approaches to address it. Section 2.4 extends the registration approach to multiple trajectories.

2.1 Temporal Registration

Let X and Y be two time-series of length n_x and n_y representing two 3-dimensional trajectory instances of the same gesture: $X \in \mathbb{R}^{3,n_x}$ and $X_t \in \mathbb{R}^3$ is a trajectory point for $1 \leq t \leq n_x$. In this section, we assume that one time-series can be derived from the

other by a time-warping $\phi = (\phi^x, \phi^y, m)$, modulo additional small spatial variations that are for instance due to style or environment specificities. The warping ϕ models a temporal alignment between the time-series X and Y, using a common timeline of length m: X_i is aligned with Y_j if and only if there exists $1 \leq t \leq m$ such that $\phi_t^x = i$ and $\phi^y = j$. The warping verifies the constraints: $0 \leq \phi_t^a - \phi_{t-1}^a \leq 1$ (*continuity*), $\phi_1^a = 1$, $\phi_m^a = n_a$ (*boundaries*), for $a \in \{x, y\}$. DTW [6] is an algorithm based on dynamic programming that optimally computes a warping ϕ minimizing the cost:

$$C_{warping}(\phi) = \sum_{t=1}^{m} \|X_{\phi_t^x} - Y_{\phi_t^y}\|^2. \tag{1}$$

An example of alignment is shown in fig. 1(d). If the trajectories are spatially very different, this approach fails. An alternative is to express the total cost in terms of features invariant to certain spatial transformations. Instead, we propose to jointly integrate the spatial registration in the process, as will be needed for template generation.

2.2 Rigid Spatial Registration

Let $\{U_i\}_{1 \leq i \leq n_u}$ and $\{V_j\}_{1 \leq j \leq n_v}$ be two 3D point sets describing a similar shape. Assuming that p point correspondences $\{(U_{\psi^u(l)}, V_{\psi^v(l)}) \mid 1 \leq l \leq p\}$ are known, where $\psi : \{1, \ldots, p\} \longrightarrow \{1, \ldots, n_u\} \times \{1, \ldots, n_v\}$, [7] optimally recovers the rigid spatial transformation $[R, T] \in SO^3 \times \mathbb{R}^3$ between the two shapes using the singular value decomposition (SVD) to minimize the sum of squared errors:

$$C_{rigid}(R, T) = \sum_{l=1}^{p} \|RU_{\psi^u(l)} + T - V_{\psi^v(l)}\|^2. \tag{2}$$

Note that the scaling factor is omitted, as the gestures we analyze interact with an environment of constant size and we do not want an additional scaling transformation.

In the case where point correspondences are not known, the iterative closest point (ICP) algorithm has been proposed [8] to iteratively estimate correspondences and the spatial registration. This approach can fail however if the initial shapes are far from the solution and the point clouds do not describe neatly the shape, as it occurs for a trajectory that overlaps in a small spatial volume. Trajectories are not simple point clouds but have a temporal dimension. In the next section, we will use this additional information to improve point assignments and thereby the registration.

2.3 Spatio-Temporal Time Warping (STW)

Let X and Y be two time-series representing two 3-dimensional trajectory instances of the same gesture, as in section 2.1. We make here the less restrictive assumption that one time-series can be derived from the other by a rigid spatial transformation $[R, T]$ and a time-warping $\phi = (\phi^x, \phi^y, m)$, modulo additional small variations. Our objective is to recover jointly the spatial transformation and the time-warping, in order to isolate and compare these variations. We propose to minimize the following registration cost

$$C_{stw}(\theta) = \sum_{t=1}^{m} \|RX_{\phi_t^x} + T - Y_{\phi_t^y}\|^2, \tag{3}$$

Algorithm 1. Spatio-temporal time warping (STW)

input : X,Y
output: $\theta = \{R, T, \phi\}$
initialization: $R = \mathbf{Id}$, $T = \mathbf{0}$;
repeat
  Compute temporal warping ϕ between $(RX + T)$ and Y using DTW;
  Compute $[R, T]$ using rigid registration and correspondence set $\{(X_{\phi_t^x}, Y_{\phi_t^y}) \mid 1 \leq t \leq m\}$ between trajectories X and Y;
until C_{stw} *converges* ;

with respect to the parameters $\theta = \{R, T, \phi\}$. We use a coordinate descent approach similar to [9], in which each step monotonically decreases the cost. We iteratively compute $[R, T]$ and ϕ using rigid registration and DTW. Even though convergence is only guaranteed to a local minimum, experiments show that the good correspondences obtained by temporal registration permit to recover large spatial transformations, even with trajectories containing variations. The approach is given in Algorithm 1.

2.4 Spatio-Temporal Registration of Multiple Trajectories

Let us now assume $X^1, \ldots, X^K$ to be K time-series of length $n_1, \ldots, n_K$, representing 3-dimensional trajectory instances of the same gesture. We make the assumption that all instances can be derived from a gesture template Y of length n_y by rigid spatial transformations $[R^k, T^k]$ and time-warpings $\phi^k = (\phi^{k,x}, \phi^{k,y}, m_k)$, modulo additional small variations due to style or environment specificities (see e.g. fig. 3(a)). Our objective is to recover all spatial transformations and time-warpings, in order to isolate and compare jointly these variations. We therefore introduce the following registration cost

$$C_{multi}(\omega) = \sum_{k=1}^{K} \sum_{t=1}^{m_k} \| R^k X^k_{\phi_t^{k,x}} + T^k - Y_{\phi_t^{k,y}} \|^2, \tag{4}$$

to be optimized on the set of parameters $\omega = \{R_k, T_k, \phi^k, Y \mid 1 \leq k \leq K\}$. During the optimization, we iteratively compute the warpings $\{\phi^k\}$, the template Y and the transformations $\{[R^k, T^k]\}$. If we suppose Y to be known, for instance to be one of the X^k, similarly to Algorithm 1 we can compute all $\{\phi^k\}$ and then all $\{[R^k, T^k]\}$ alternatively, to decrease monotonously $C_{multi}(\theta)$. The template needs however to be refined, as a particular sequence may contain anomalies or not be representative of the timeline. To compute the template Y, we use a temporal averaging technique presented in [10]. This approach synchronizes several time series to a common timeline whose length is the average length of all time series, using an initial template. Using this synchronization, the time-series are then averaged. This method is presented in Algorithm 2, where the monotonicity of the $\tilde{\phi}^{k,x}$ guaranties the invertibility of μ. Interpolation steps are needed in the computations of μ and Z, but omitted for better readability.

The final registration method is presented in Algorithm 3. Since the template computation also modifies the temporal warpings, monotonous decrease of the cost is not obvious. Convergence is however verified in our experiments.

Algorithm 2. Template generation

input : $\{X^k\}_{1\leq k\leq K}$, initial template $\tilde{Y}$
output: new template Y and new warpings ϕ^k
for $1 \leq k \leq K$ **do**
  Compute warpings $\tilde{\phi}^k = (\tilde{\phi}^{k,x}, \tilde{\phi}^{k,y}, \tilde{m})$ between X^k and $\tilde{Y}$ using DTW;
Compute $\mu(t) = \sum_{k=1}^{K} \tilde{\phi}_t^{k,x}$, where $1 \leq t \leq \tilde{m}$;
Define $Y_t = \sum_{k=1}^{K} \tilde{Y}_{(\tilde{\phi}^{k,x}(\mu^{-1}(t)))}$, where $1 \leq t \leq \frac{1}{K}\sum_{k=1}^{K} n_k$;
for $1 \leq k \leq K$ **do**
  Compute warpings $\phi^k = (\phi^{k,x}, \phi^{k,y}, m)$ between X^k and Y using DTW;

Algorithm 3. Multiple trajectories registration

input : $\{X^k\}_{1\leq k\leq K}$
output: $\omega = \{R_k, T_k, \phi^k, Y \mid 1 \leq k \leq K\}$
Y is initialized as the X^k with median length;
for $1 \leq k \leq K$ **do**
  $R^k = \mathbf{Id}, T^k = \mathbf{0}$
repeat
  Update Y and $\{\phi^k\}$ using Algo. 2 with input $\{R^k X^k + T^k\}_{1\leq k\leq K}$ and Y;
  for $1 \leq k \leq K$ **do**
    compute R^k, T^k using rigid registration and correspondence set $\{(X^k_{\phi_t^{k,x}}, Y_{\phi_t^{k,y}}) \mid 1 \leq t \leq m_k\}$;
until C_{multi} *converges* ;

3 Experiments and Results

3.1 Synthetic Experiments

Pair of trajectories
Let X and Y be trajectories describing 80% of a circle (fig. 1(a)). Y is obtained by rotating X in the same plane by 60 degrees and translating it. A noise of 5% is added to the data points. The trajectories are performed at different speeds, as illustrated by the ground truth temporal warping in fig. 1(d). Y mostly describes the circle twice as fast, but in the end at the same speed ($n_x = 150$ and $n_y = 88$). Fig. 1(b) shows the results for ICP, which fails due to the noise and spatial overlap. DTW alone also fails (fig. 1(e,f)), as it assumes a time warping at the extremities and tends to associate temporally the points closest spatially. CTW[9] gives better results (fig. 1(g,h)), but also fails to recover the temporal alignment since it does not model the appropriate transformation. The affine deformation of the data is clearly visible in fig. 1(g). Finally, the STW approach recovers the correct spatio-temporal warping (fig. 1(i,j)).
Multiple trajectories
Figure 2 illustrates the registration of multiple trajectories using the approach of section 2.4. Three noisy partial circles with different time-warpings (lengths 88,150,250) are correctly registered, as shown in fig. 2(b). Note that the shape of the template is also a circle, which would not be the case if the rotations were not properly recovered.

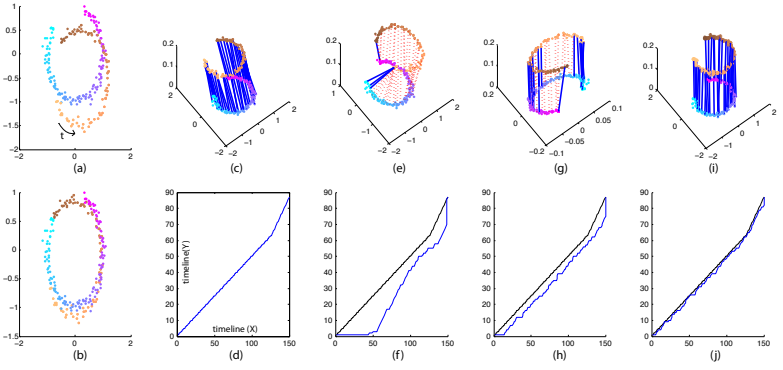


Fig. 1. Synthetic experiment. (a) two trajectories; (b) spatial registration using ICP; (c) Ground truth spatio-temporal registration, with trajectories in two different planes for better visualization. Lines indicate the spatial alignment; (d) corresponding time warping; (e) registration using DTW only; (g) registration using CTW; (i) registration using STW; (f),(h),(j) show the time warpings corresponding to (e),(g),(i). The dotted red lines in (e),(g),(i) indicate wrong alignment, with tolerance ± 5 frames.

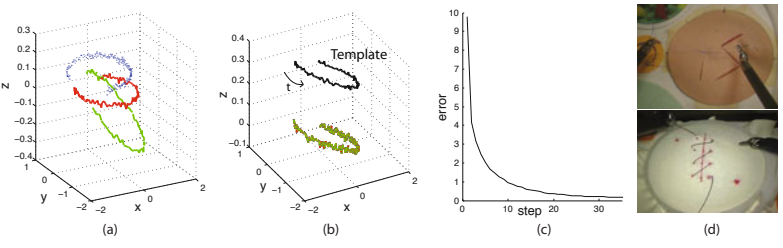


Fig. 2. Joint registration of 3 synthetic trajectories / setups for real experiments. (a) raw trajectories; (b) registration with generated template isolated on top for better visualization; (c) registration error; (d) setups for knot-tying (top) and 4-throw suture (bottom).

3.2 Experiments on Tele-Manipulation Data

We are interested in temporally synchronizing and spatially superimposing trajectories, to replay synchronously the videos or the registered gestures in a virtual environment.

Knot-tying analysis

We use six sequences of a knot tying task performed by the same user using a da Vinci robot (fig. 2(d)). Trajectories describe the left tool positions. In each recording, the task pod has a different position and orientation. In spite of the large spatial transformations and the multiple local variations (fig. 3(a)), e.g. in directions when pulling the thread, all trajectories can be jointly registered. Fig. 3(b) shows the qualitative result and the resulting template. Fig. 3(c) shows the decrease of the registration error C_{multi} during the optimization. The temporal synchronization between two trajectories is shown in

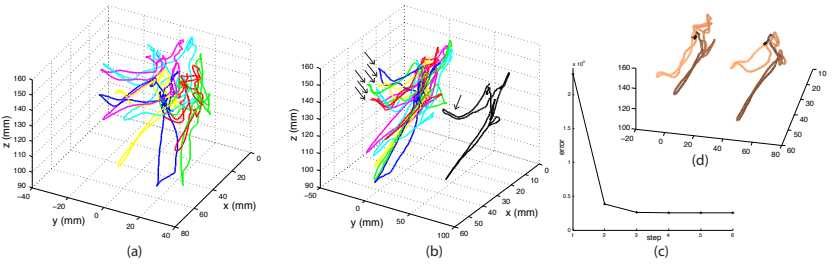


Fig. 3. Knot tying. (a) 6 raw trajectories; (b) registration with template isolated on the right for better visualization; arrows show a registered feature; (c) registration error; (d) temporal synchronization between two trajectories pictured by a color gradient.

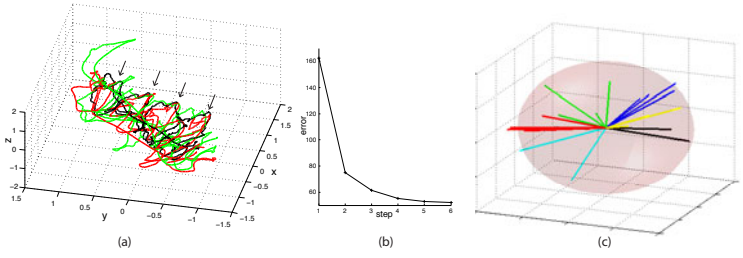


Fig. 4. Four-throw suture. (a) 3 of 19 registered trajectories with arrows showing the registered throws; (b) registration error; (c) rotation axis, colored per surgeon.

fig. 3(d) using a color gradient to represent time. By registering the sequences all together, the performances of surgeons can be synchronously compared (e.g. beginners vs experts). Side by side virtual replay can also teach a trainee the correct tool orientation.

Four-throw suture analysis

We use 19 sequences of a 4-throw suture performed by 6 different surgeons using the da Vinci tele-surgical robot (see fig. 2(d)). The trajectory data describes the cartesian positions of the master-manipulator corresponding to the right hand and directly controlling the right instrument. Fig. 4(a) shows a subset of the registered sequences for qualitative evaluation. Fig. 4(b) shows the quantitative registration error C_{multi}. This dataset illustrates another interesting application of the registration. For each surgeon, the same suturing pod is used, but different surgeons use different pods. Even though the data was meant to be recorded with the same setup, the videos show that between surgeons, the pods are rarely at the same location, but rotated. Fig. 4(c) shows the computed 3D axis of the axis/angle representations of the rotations R^k, colored per surgeon. We see that the registration is precise enough to cluster the 3D axis per surgeons (and setups), in spite of the fact that the long data sequences contain multiple atomic gestures and variations.

4 Discussion and Conclusion

Trajectories occur naturally in computer assisted surgery: they describe the movements of tools, markers and body features. We first propose an approach to jointly register spatially and temporally two trajectories differing by a rigid spatial transformation and a temporal warping, and containing possibly additional local variations. Such an assumption is valid for most complex gestures, since both the task to accomplish and the environment constrain the operator's freedom of movement. We then state the registration problem for multiple trajectories and propose an optimization approach that simultaneously generates a representative template. Synthetic experiments show the usefulness of combining spatial and temporal registration. Real experiments demonstrate its applicability for registering complex gestures, like knot tying and multi-throw suturing, for example for intelligent side-by-side comparison. Even though we register multiple trajectories of a single object (a tool), by stacking up the data, one can extend the approach to register multiple trajectories of multiple objects. We are interested in applying this idea to the registration of medical images acquired over time.

Acknowledgments. This work is funded by NSF grant CDI-0941362. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

1. Blum, T., Sielhorst, T., Navab, N.: Advanced Augmented Reality Feedback for Teaching 3D Tool Manipulation, ch. 25, pp. 223–236. Lupiensis Biomedical Publications (2007)
2. Lin, H.C., Shafran, I., Murphy, T.E., Okamura, A.M., Yuh, D.D., Hager, G.D.: Automatic detection and segmentation of robot-assisted surgical motions. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3749, pp. 802–810. Springer, Heidelberg (2005)
3. Varadarajan, B., Reiley, C., Lin, H., Khudanpur, S., Hager, G.: Data-derived models for segmentation with application to surgical assessment and training. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5761, pp. 426–434. Springer, Heidelberg (2009)
4. Padoy, N., Blum, T., Ahmadi, A., Feussner, H., Berger, M.-O., Navab, N.: Statistical modeling and recognition of surgical workflow. Medical Image Analysis (2010), doi:10.1016/j.media.2010.10.001
5. van den Berg, J., Miller, S., Duckworth, D., Hu, H., Wan, A., Fu, X.-Y., Goldberg, K., Abbeel, P.: Superhuman performance of surgical tasks by robots using iterative learning from human-guided demonstrations. In: ICRA 2010, pp. 2074–2081 (2010)
6. Sakoe, H., Chiba, S.: Dynamic programming algorithm optimization for spoken word recognition. IEEE Trans. Acoust. Speech Signal Process. 26(1), 43–49 (1978)
7. Umeyama, S.: Least-squares estimation of transformation parameters between two point patterns. IEEE Trans. Pattern Anal. Mach. Intell. 13, 376–380 (1991)
8. Zhang, Z.: Iterative point matching for registration of free-form curves and surfaces. Int. J. Comput. Vision 13, 119–152 (1994)
9. Zhou, F., De la Torre, F.: Canonical time warping for alignment of human behavior. In: Advances in Neural Information Processing Systems Conference, NIPS (2009)
10. Wang, K., Gasser, T.: Alignment of curves by dynamic time warping. Annals of Statistics 25(3), 1251–1276 (1997)

A Non-disruptive Technology for Robust 3D Tool Tracking for Ultrasound-Guided Interventions

Jay Mung[1], Francois Vignon[2], and Ameet Jain[2]

[1] University of Southern California
[2] Philips Research North America

Abstract. In the past decade ultrasound (US) has become the preferred modality for a number of interventional procedures, offering excellent soft tissue visualization. The main limitation however is limited visualization of surgical tools. A new method is proposed for robust 3D tracking and US image enhancement of surgical tools under US guidance. Small US sensors are mounted on existing surgical tools. As the imager emits acoustic energy, the electrical signal from the sensor is analyzed to reconstruct its 3D coordinates. These coordinates can then be used for 3D surgical navigation, similar to current day tracking systems. A system with real-time 3D tool tracking and image enhancement was implemented on a commercial ultrasound scanner and 3D probe. Extensive water tank experiments with a tracked 0.2mm sensor show robust performance in a wide range of imaging conditions and tool position/orientations. The 3D tracking accuracy was 0.36 ± 0.16mm throughout the imaging volume of $55^\circ \times 27^\circ \times 150$mm. Additionally, the tool was successfully tracked inside a beating heart phantom. This paper proposes an image enhancement and tool tracking technology with sub-mm accuracy for US-guided interventions. The technology is non-disruptive, both in terms of existing clinical workflow and commercial considerations, showing promise for large scale clinical impact.

Keywords: Ultrasound, Tool Tracking, Interventions, Navigation.

1 Introduction

X-ray and Ultrasound (US) are ubiquitous for guidance of interventional procedures, with the latter being used more in terms of total procedure numbers. Estimated at over 10 million procedures per year, US-guided interventions range from routine needle insertion for regional anesthesia [1] to biopsies [2] and percutaneous ablation of cancer [3] to more advanced procedures such as structural heart repair [4]. Its low cost, ease of use, excellent soft tissue visualization and use of non-ionizing radiation make it a popular modality for both diagnostics and interventions. The biggest limitation however is its lack of robust tool visualization. X-ray, despite its lack of soft tissue contrast and harmful radiation, has hence been the primary modality to fill this gap. With the advent of advanced

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 153–160, 2011.

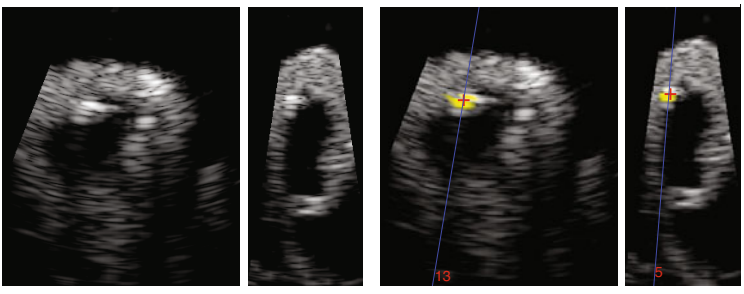


Fig. 1. Left: azimuth and elevation slices through a 3D US image of a catheter in a beating heart phantom. Note the difficulty in tool detection. Right: same, with tool enhancement enabled. Blue lines indicate positions of orthogonal slices. Note that the "unenhanced" images are automatically selected to display the catheter tip.

surgical navigation and image-fusion, research attempts have been made to improve tool visualization. However, these attempts have not yet fully translated to large-scale clinical practice. In particular, external tracking technologies have made some progress at the expense of clinical workflow, leading to slow adoption rates. Clinically viable and robust 3D tool visualization still remains a challenge.

The "invisible tool" phenomenon is well known in US imaging and interventions (Figure 1 Left). Several approaches have been proposed to try to overcome this problem. These include roughening the tool surface to make it more isotropically reflective [5]. This has translated clinically [6], but the solution is not yet robust enough for difficult interventions [5]. Ultrasound beamforming protocols have also been modified to better detect straight specular reflectors [7]. This also has translated clinically [8]. However such a solution currently works only for a limited category of ultrasound probes (linear or curvilinear geometry), limited needle incidence angles with some a priori knowledge of needle orientation, and inside sub-regions in the tissue image. Doppler imaging of a vibrating tool has also been attempted to enhance tool visualization [9], however requiring changes to the imaging pipeline including a drop in imaging frame rate and interference from moving tissue. Note that all aforementioned technologies only provide image enhancement of the tool but do not yield absolute position information of the tool. External tracking technologies such as optical tracking have been introduced as solutions providing absolute instrument tracking [10]. Since these methods alter the existing clinical workflow, they have seen a limited clinical translation, mostly in orthopedic and neurological interventions. A robust solution for US guided interventions is still needed.

A promising direction is to mount a small US "transponder" element on the tool that, upon receiving ultrasound energy from the US beams sent from the imaging probe, re-radiates ultrasound energy toward the probe, thus acting as a very bright reflector. The concept was first introduced in 1984 [11]. At that time, a passive receiver method was also introduced where upon reception, instead of re-radiating ultrasound energy, a low-voltage electrical signal is passed to

the system for image enhancement. This method was introduced clinically by ATL to enhance needle guidance of breast biopsies [12]. Another embodiment was later described by Vilkomerson [13] and used in a similar way for aiding catheter ablation in pigs [14]. These systems provided enhanced tool visualization but not position and orientation coordinates of the tracked tools. Furthermore these early accounts made use of 2D US imaging technology which limited the enhancement feature to a plane. For these reasons, they had limited clinical utility. Building upon this technology, an offline technique was later proposed to detect location by analyzing the ultrasound signals received at the tracked transducer from an experimental 3D imaging probe [15]. Though promising, this technique was not integrated with an imaging system for image enhancement and interventional guidance. Alternately, attempts have been made to localize intra-body ultrasound transducers by deploying a number of discrete ultrasound elements to track the tranducer's position via trilateration [16,17]. However, these still require a registration of tool position to the US image, making the method difficult to translate clinically.

Key Contributions: As far as the authors are aware, this paper is the first to propose a system for real-time 3D tool tracking and image enhancement on a commercial scanner. It tracks a US sensor by using the US signals impinging on it to compute its 3D position. The key contributions are:

- A new technique to achieve 3D tool tracking with an accuracy of 0.36 mm
- Intuitive tool image enhancement with automatic 2D slice selection (Fig. 1)
- Extensive validation of a prototype integrated with a commercial scanner
- A solution with potential for clinical implementation

2 Methods

Principle of Operation: US sensor(s) are embedded at known locations on the surgical tool. For position tracking, angular position is derived by knowing the angular direction(s) of the US beam(s) that impinge on the receiver, while depth information is derived from the time-of-flight from emission of the respective beam(s). Tracking two receivers can yield orientation of the tool. For image enhancement, (i) an image of the received signals as a function of depth and angle; or (ii) the 3D coordinate can be overlaid on the US image. A schematic description of the principle of operation is provided in Fig. 2a.

Tracking: A 3D US imaging probe emits ultrasound beams that regularly sample the field of view (FOV) in an array of $N_{AZ} \times N_{EL}$ beams in the azimuth and elevation directions, respectively (the geometry of US acquisition is depicted in Fig. 2b). The temporal signals sensed by the receiver during the acquisition of one image are formatted in a 3D $N_{AZ} \times N_{EL} \times time$ "data matrix". In order to retrieve angular position information with a better resolution than that given by the spacing of the US beams, first a maximum intensity projection (MIP) of the

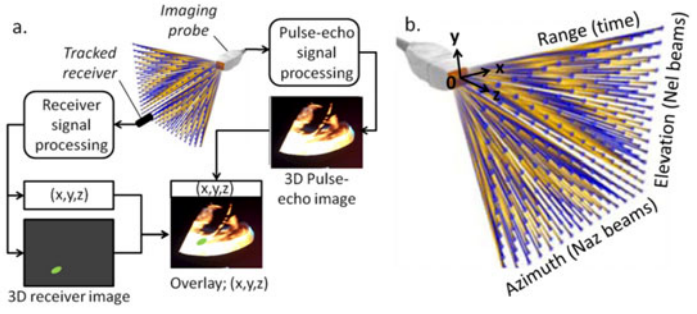


Fig. 2. a: Schematic of principle of operation. b: Depiction of the coordinate systems.

"data matrix" over the time dimension is performed to yield a 2D ($N_{AZ} \times N_{EL}$) MIP matrix on which a 2D Gaussian fit is applied. The Gaussian center is used to estimate the angular coordinates of the receiver in the US coordinate system. The depth information is obtained by finding the time at which the maximum signal arrives at that angle and multiplying by the speed of sound. Thus with angle and range information, the 3D sensor location is computed.

Display Enhancement: The computed position of the sensor is used to slice through the 3D pulse-echo data received by the scanner and the 3D image of the receiver, extracting two orthogonal 2D slices through the sensor - one at constant azimuth and the other at constant elevation angle. These two slices are displayed as an overlay of sensor position (in color) on the standard pulse-echo US image (in grayscale).

Experimental Setup: The 3D imaging X7-2 probe is driven by an iE33 scanner (Philips Healthcare, Bothell, WA) in "Live 3D" mode with a FOV of $\approx 55°(azimuth) \times 27°(elevation) \times 150mm(range)$ at an imaging rate of 21Hz. For this application, the iE33 is equipped with an in-house system that streams the scanner's radiofrequency pulse-echo data to a separate workstation at a lowered frame rate. The US receiver is a PVDF hydrophone with a 0.2mm aperture equipped with a 20dB pre-amplifier (model HGL-0200, Onda Corp., Sunnyvale, CA). Sensor acquisitions are synchronized by the iE33 scanner's 'beam out' and 'volume out' trigger signals that mark the start of emission of each US beam and each US volume (an ensemble of $N_{AZ} \times N_{EL}$ beams), respectively. Sensor data is streamed to a workstation at the same rate of the scanner's pulse-echo RF data through a data acquisition card (Gage Applied Technologies, Lachine, Canada). The receiver is suspended in a water bath by a 3D motorized stage (model MM4006, Newport Corp., Irvine, CA) that is controlled by the central workstation through a GPIB interface. The stage is specified accurate and repeatable to <5 μm and thus provides ground-truth coordinates. A manual rotation component is also available in the x-z plane. Streamed data from the scanner and sensor are processed live using Matlab to yield tracked sensor positions and overlay volume 2D slices with a prototype frame rate of 0.26Hz.

3 Experiments and Results

System Calibration: The first step prior to the validation experiments was to calibrate the system. Data was acquired with the stage translating the receiver in 1 mm increments over four rectangles in x-y parallel planes at depths of 20mm, 60mm, 100mm, and 130mm. A total of 2000 data points were collected. An unconstrained nonlinear optimization minimizing fiducial registration error calibrated the following variables: speed of sound (SOS) in water, transmit beam angles in water, and time offset between the imaging system 'beam out' trigger and US probe beam emission. The calibrated SOS was 1491.6 m/s. These parameters were used for the subsequent validation experiments.

3.1 Tracking Accuracy

Tracking Accuracy Over the 3D US FOV: The first experiment evaluated the system accuracy throughout the extent of the 3D US FOV. The system acquired data in the same manner as the calibration step (Fig. 3), collecting a total of 5110 data points. 10% of the points were chosen at random to calculate a registration matrix using Procrustes analysis. The registration matrix was then applied to the remaining 90% of the points, from which Target Registration Error (TRE) values were calculated. The 10% random point selection and registration was repeated 100 times to provide a total of 459,900 TRE values. The TRE values were then grouped by slice depth and azimuth angle, with the results shown in Fig. 4. The accuracy is 0.36 ± 0.16mm across the FOV with no significant biases in any one direction. Furthermore, there is no significant correlation to the azimuth angle. The 2.1% distance error at 20mm separation suggests a 1.2° mean rotation accuracy using two sensors separated by that distance.

Accuracy as a Function of Depth: A second experiment was performed to evaluate the system accuracy throughout depth at the center of the US FOV. The experimental protocol is identical to the first experiment except that the

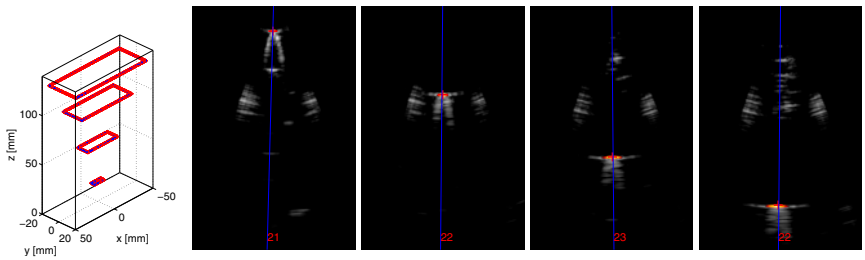


Fig. 3. Left: the tracked sensor ground truth and estimated positions (red crosses and blue stars, respectively). Right: sample overlay images ('hot' colormap receiver images with grayscale pulse-echo images) at the four rectangle depths (20mm, 60mm, 100mm, 130mm). Shown are 4 automatically selected azimuth slices through 3D datasets. The resolution degrades with increasing depth.

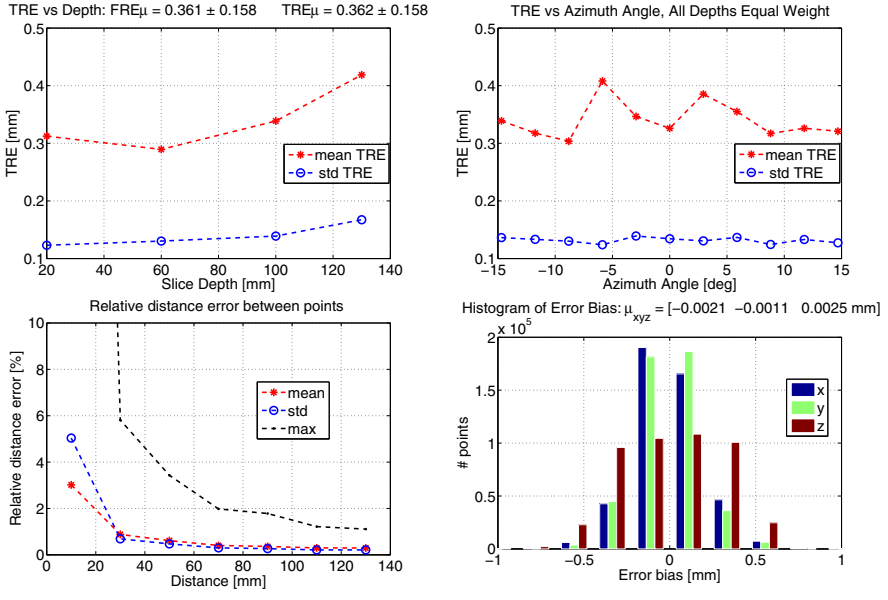


Fig. 4. Error measures from the first validation experiment: Top left: accuracy vs. depth. Top right: accuracy vs. azimuth angle. Bottom left: Percent distance error vs. distance between two points. Bottom right. Histograms of error bias values in X, Y, Z.

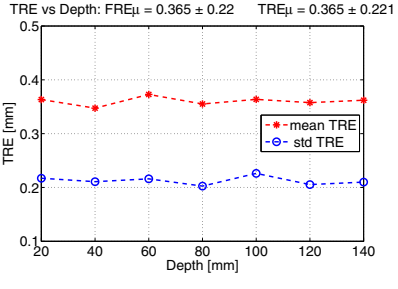


Fig. 5. Error measures from z-translation validation experiment: Accuracy vs. z-depth

stage translated the receiver in 1 mm increments along the center z-axis from 10 to 140mm for a total of 141 points per run. A total of 5110 data points were collected, representing 36 runs through the straight line. The system accuracy was evaluated with the same routine as the first experiment. The accuracy values are grouped by depth and shown in Fig. 5. The accuracy remains constant at 0.37mm across all depths. Note that the average performance is same as before, but depth dependent variation is low due to constant sensor orientation.

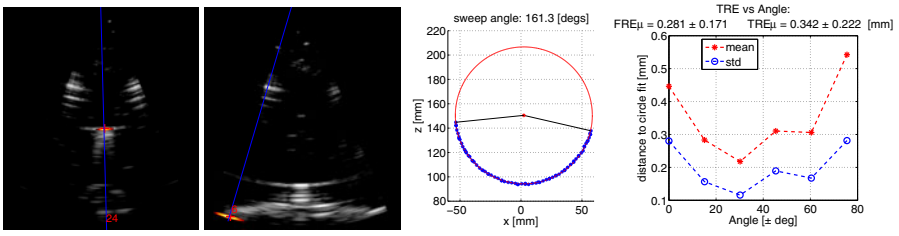


Fig. 6. Left: enhanced images of the hydrophone at approximately 0° and 90° incidence. Center: Recorded positions with circle fit. Right: TRE vs. hydrophone angle.

3.2 Sensor Orientation

A third experiment evaluated the system accuracy as a function of sensor angle. Here the receiver was mounted on a manual rotational stage with the center of rotation approximately 150mm from the probe surface. The receiver tip was held at a radius of about 55mm from the center of rotation and manually rotated from approximately −80° to 80° for a total 73 data points. The collected data points were divided into 6 bins according to angle. Each bin of data was then selectively omitted so that 5 bins would be used for a circle-fit, while the omitted bin provided TRE values (distance to the circle). The accuracy is shown in Fig. 6, showing that the technique is robust to sensor orientations (<0.6mm error).

3.3 Beating Heart Phantom

The fourth and final experiment tested the ability to dynamically track a catheter prototype consisting of a 3mm diameter by 5mm long PZT cylinder sensor mounted on a 1.5m thin-gauge coaxial cable [16]. It was tested with an in-house beating heart phantom, a polymer replica with heart-like motion and US imaging appearance. Note in figure 1 the difficulty of tool detection without the image enhancement and tool tracking enabled. The received waveform was elongated at some angles likely due to the sensor's bandwidth and spatial sensitivity pattern. This had a deleterious effect on the overlay resolution (Fig. 1 right). Overall, this experiment qualitatively illustrates the added value of the tracking in a more realistic situation, with some amount of ultrasound aberration and absorption, and the presence of motion.

4 Conclusion and Future Work

We demonstrated real time US-based tool tracking on an FDA approved commercial scanner. Experiments indicate an accuracy of 0.36 ± 0.16 mm and show robustness to varying imaging conditions. Because image enhancement occurs in US imaging space, the tool location and US image are always perfectly registered even in the face of image distortion. This method is non-disruptive where US is already part of the clinical workflow and requires no additional major capital equipment, thereby facilitating clinical adoption. However, this method requires that interventional tools incorporate US sensors and electrical connections safely

and unobtrusively. This is an engineering task for future work. Future work also includes improving frame rate, pre-clinical and clinical validation and determining which US-guided procedures would benefit most from this technology.

References

1. Chin, K.J., Perlas, A., Chan, V.W., Brull, R.: Needle Visualization in Ultrasound-Guided Regional Anesthesia: Challenges and Solutions. Reg. Anesth. Pain. Med. 33(6), 532–544 (2008)
2. O'Flynn, E.A., Wilson, A.R., Michell, M.J., et al.: Image-Guided Breast Biopsy: State-of-the-Art. Clin. Radiol. 65(4), 259–270 (2010)
3. Liveraghi, T.: Radiofrequency Ablation of Hepatocellular Carcinoma. Surg. Oncol. N. Am. 20(2), 281–299 (2011)
4. Balzer, R., Kelm, M., Kühl, H.P.: Real-Time Three-Dimensional Transoesophageal Echocardiography for Guidance of non-Coronary Interventions in the Catheter Laboratory. Eur. J. Echocardiogr. 10(3), 341–349 (2009)
5. Nichols, K., Wright, L.B., Spencer, T., Culp, W.C.: Changes in Ultrasonographic Echogenicity and Visibility of Needles with Changes in Angles of Insonation. J. Vasc. Interv. Radiol. 14(12), 1553–1557 (2003)
6. EchoTip®, Cook Medical, `http://www.cookmedical.com/wh/dataSheet.do?id=4480`
7. Cheung, S., Rohling, R.: Enhancement of Needle Visibility in Ultrasound-Guided Percutaneous Procedures. Ultr. Med. Biol. 30(5), 617–624 (2004)
8. Sonosite Needle Visualization, `http://www.sonosite.com/needlevisualization/`
9. Idriss, S.F., Fronheiser, M.P.: Vibrating Interventional Device Detection Using Real-Time 3-D Color Doppler. IEEE Trans. Ultr. Ferr. Freq. Contr. 55(6), 1355–1362 (2008)
10. IGS Platforms, `www.brainlab.com/art/2798/4/image-guided-surgery-platforms/`
11. Breyer, B., Cikes, I.: Ultrasonically marked catheter-a method for positive echographic catheter position identification. Med. Biol. Eng. Comp. 22(3), 268–271 (1984)
12. Perrella, R.R., Kimme-Smith, C., Tessler, F.N., Ragavendra, N., Grant, E.G.: A new electronically enhanced biopsy system: value in improving needle-tip visibility during sonographically guided procedures. Am. J. Roentgenol. 158(1) (1992)
13. Vilkomerson, D., Lyons, D.: A system for ultrasonic beacon-guidance of catheters and other minimally-invasive medical devices. IEEE Trans. UFFC 44(2), 496–504 (1997)
14. Menz, V., Vilkomerson, D., Ren, J.F., Michele, J.J., Schwartzman, D.: Echocardiographic Transponder-Guided Catheter Ablation Feasibility and Accuracy. J. Interv. Card. Electrophysiol. 5(2), 203–209 (2001)
15. Merdes, C.L., Wolf, P.D.: Locating a catheter transducer in a three-dimensional ultrasound imaging field. IEEE Trans. Biomed. Eng. 48(12), 1444–1452 (2001)
16. Mung, J., Han, S., Yen, J.T., et al.: Design and in vitro evaluation of a real-time catheter localization system using time of flight measurements from seven 3.5 MHz single element ultrasound transducers towards abdominal aortic aneurysm procedures. Ultrasonics 51(6), 768–775 (2011)
17. Meyer, S.A., Wolf, P.D., et al.: Application of Sonomicrometry and Multidimensional Scaling to Cardiac Catheter Tracking. IEEE Trans. on Biomed. Eng. 44(11), 1061–1067 (1997)

Image-Based Device Tracking for the Co-registration of Angiography and Intravascular Ultrasound Images

Peng Wang[1], Terrence Chen[1], Olivier Ecabert[2], Simone Prummer[2], Martin Ostermeier[2], and Dorin Comaniciu[1]

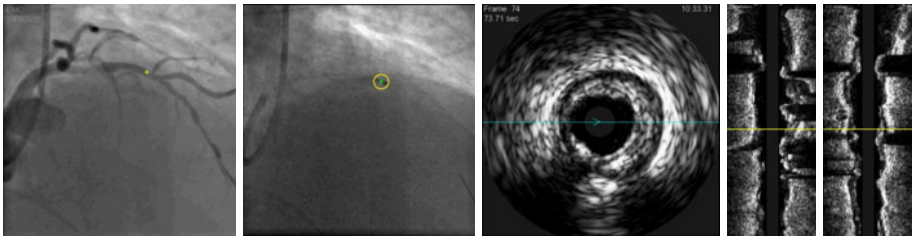
[1] Siemens Corporate Research, Siemens Corporation, Princeton NJ, U.S.A
[2] Siemens AG, Healthcare Sector, Siemensstr. 1, Forchheim, Germany

Abstract. The accurate and robust tracking of catheters and transducers employed during image-guided coronary intervention is critical to improve the clinical workflow and procedure outcome. Image-based device detection and tracking methods are preferred due to the straightforward integration into existing medical equipments. In this paper, we present a novel computational framework for image-based device detection and tracking applied to the co-registration of angiography and intravascular ultrasound (IVUS), two modalities commonly used in interventional cardiology. The proposed system includes learning-based detections, model-based tracking, and registration using the geodesic distance. The system receives as input the selection of the coronary branch under investigation in a reference angiography image. During the subsequent pullback of the IVUS transducers, the system automatically tracks the position of the medical devices, including the IVUS transducers and guiding catheter tips, under fluoroscopy imaging. The localization of IVUS transducers and guiding catheter tips is used to continuously associate an IVUS imaging plane to the vessel branch under investigation. We validated the system on a set of 65 clinical cases, with high accuracy (mean errors less than 1.5mm) and robustness (98.46% success rate). To our knowledge, this is the first reported system able to automatically establish a robust correspondence between the angiography and IVUS images, thus providing clinicians with a comprehensive view of the coronaries.

1 Introduction

In image guided interventions, accurate and robust localization of medical devices can provide valuable information to improve clinical workflow and to facilitate the operations in interventions. In this paper, we present a computational framework for medical device detection and tracking in X-ray images, and demonstrate it with a novel application for the co-registration of 2D angiography and intra-vascular ultrasound (IVUS) images. The framework is briefly illustrated in Fig. 1. The X-ray angiography is the primary modality that guides percutaneous interventions (PCI), providing the spatial structure information of coronaries. The IVUS images can provide rich information on the vessel wall

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 161–168, 2011.



(a) (b) (c) (d)

Fig. 1. Device tracking for the co-registration of IVUS and angiography images. (a): a angiogram frame. The yellow cross indicate a registered IVUS imaging plane; (b): a fluoroscopic frame. The yellow circle represents a tracked IVUS transducer; (c)(d): cross-sectional and axial view of IVUS images. The yellow lines in the axial views indicate the registered IVUS image.

composition and extent of plaques, being used for tissue characterization, the analysis of lesion, and the estimation of stent expansion. However, the lack of the vessel orientation information makes it difficult to fully understand the spatial structure of vessels. The co-registration of the two image modalities will combine the high spatial resolution from the IVUS data with the good overview and 2D/3D orientation obtained from angiography, facilitating the usage of IVUS in cathlab and improving clinical outcome of cardiac interventions [1,2].

To register the two image modalities, we need to find the position of IVUS image plane along a vessel branch during the IVUS pullback. Manual labeling, as adapted in few preliminary systems [1,3], is straightforward. But manually finding the correspondences between IVUS and angiography data needs expertise and can be time-consuming. Furthermore, without continuous tracking, previous work inevitably assumes a constant IVUS pullback speed to estimate IVUS imaging planes from an interpolation [3,5]. This paper presents a computational framework for automated image-based IVUS transducer detection and tracking in X-ray images to continuously identify the location of IVUS image plane, and to achieve an unconstrained co-registration of angiography and IVUS images.

Accurate and robust tracking in X-ray is challenging. The devices undergo cardiac motion and breathing motion during interventions. Even with ECG triggering, the breathing motion could still be large. Considering the low signal-to-noise ratio of fluoroscopy images and the presence of other surgical devices in the images, conventional tracking methods based on intensity appearance or edge-based energy would encounter difficulties. We present a framework that employs learning-based detections and model-based probabilistic tracking to address the aforementioned challenges. The learning-based detectors detect various devices present in this application, i.e., IVUS transducers, the body and tips of a guiding catheter in which the IVUS transducer is inserted, and the wire that is used to pull the IVUS transducers. A probabilistic model-based tracking framework is introduced to combine detections of different devices in a Bayesian inference framework, to achieve robust and accurate tracking. The device tracking results

are then used for a geodesic distance based registration to associate an IVUS imaging plane at each time stamp during pullback with a point on the coronary branch under investigation. The detection and tracking is automated, only requiring the selection of a coronary branch in angiography as an input. It does not assume a constant IVUS pullback speed, and can work under different projection angles. We validate the framework on a set of 65 clinical cases acquired in hospitals. Quantitative evaluation results show that the framework has achieved high tracking robustness and accuracy for the co-registration of IVUS and angiography images.

2 Workflow

As illustrated in Fig. 2, the system workflow includes two stages: the "angiogram" stage and the "pullback" stage. In the angiogram stage, contrast agents are injected to visualize vessel branches. A frame at the end-diastole (ED) phase is used to select the vessel branch under investigation where the IVUS imaging will take place, as the input to subsequent tracking and registration. A vessel detection method that requires minimal user interactions [7] is applied to segment the center line of a vessel branch and to extend the centerline toward the guiding catheter. In the pullback stage, an IVUS transducer is pulled back from the distal end of the target vessel branch toward the proximal end. During the pullback, both the fluoroscopic images, which are trigged by ECG at the ED cardiac phase, and IVUS images are acquired with time stamps. Due to the existence of continuous breathing motion, an additional point other than the IVUS transducer is needed as the reference point to compensate the breathing motion. The tip of a guiding catheter, through which the IVUS transducer is inserted and pull, is selected as the reference point, as it is the most stable and distinguishable point that we can find in the fluoroscopy. In the pullback stage, our method automatically detects and tracks the movements of IVUS transducers and guiding catheter tips in fluoroscopy, and maps the tracked IVUS transducers back to the selected vessel branch. Simultaneously the IVUS images are registered to the vessel, by using the time synchronization between two image modalities.

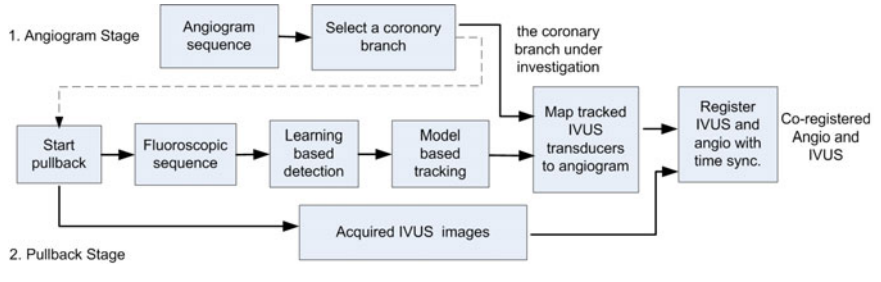
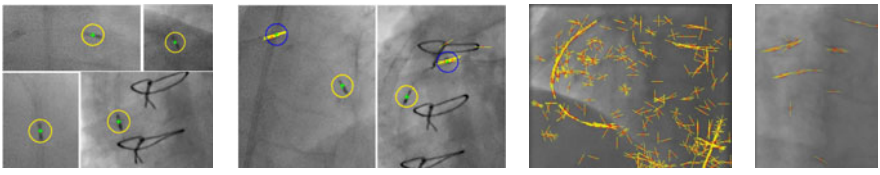


Fig. 2. The co-registration system workflow



(a) (b) (c) (d)

Fig. 3. Learning-based detection examples. (a): yellow circles represent detected IVUS transducers; (b): blue circles represent detected guiding catheter tips. Here multiple detection results close to each other are clustered into a single detection result; (c): detected segments of wire; (d) detected segments of guiding catheter body.

3 Learning-Based Device Detection

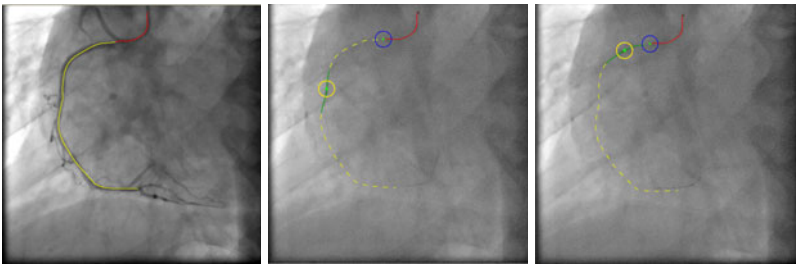
Learning-based detectors are trained from a set of off-line collected data, including both object samples (positive) and non-object samples (negative), to learn the decision boundary that separates different classes. In this work, the probabilistic boosting tree (PBT) and Haar features [4] are used to train the learning-based detectors. Here, we denote the probabilistic outputs from PBT as $P(\mathbf{Z}|\mathbf{x})$, where $\mathbf{Z}$ is the observed image, and $\mathbf{x}$ is the object state (i.e., if an image patch belongs to the positive or negative class.) Four types of detectors are trained. They are denoted as $P_{\text{IVUS\_trans}}$, $P_{\text{cath\_body}}$, $P_{\text{cath\_tip}}$ and P_{wire}, for the IVUS transducer, the guiding catheter body, the guiding catheter tip, and the wire body respectively. Note that the guiding catheter body and wire detectors aim at the detection of segments, not the whole structure. Some detection results are shown in Fig. 3. Due to image artifacts and low visibility of devices, there exist false detections. To make the tracking robust to detection errors, a tracking framework that integrates multiple detections is introduced in the Section 4.

4 Model-Based Device Tracking and Registration

4.1 A Bayesian Tracking Framework

Since the fluoroscopic images are acquired by ECG triggering, the device mainly undergoes breathing motion, which can be approximated by affine motion. The breathing motion at the t-th frame is denoted as $M_t = (m_t^x, m_t^y, m_t^r)$, where m_t^x, m_t^y, and m_t^r are the 2D translation and rotation parameters. The motion tracking is expressed as the inference of the motion parameters from the fluoroscopic sequences acquired at the pullback stage. We formalize the parameter inference in a sequential Bayesian inference framework. Assuming a commonly used Markov property for tracking, the posterior probability $P(\mathbf{M}_t = (m_t^x, m_t^y, m_t^r)|\mathbf{Z}_t)$ is given in Eqn. (1):

$$P(\mathbf{M}_t|\mathbf{Z}_t) \propto P(\mathbf{M}_t)P(\mathbf{Z}_t|\mathbf{M}_t), \tag{1}$$



(a) (b) (c)

Fig. 4. Model-based tracking. (a): a vessel detected in the angiogram phase; (b)(c): an integrated model of point and curve elements is automatically initialized by the input from the angiogram stage and continuously tracked in the pullback stage.

where $\mathbf{Z}_t$ is an observed fluoroscopic sequence. The tracking result is the motion parameter corresponding to the maximal posterior probability, i.e. $\hat{\mathbf{M}}_t = \arg\max_{\mathbf{M}_t} P(\mathbf{M}_t|\mathbf{Z}_t)$. In Eqn. (1), $P(\mathbf{M}_t)$ is the prior probability. We model the prior probability as $P(\mathbf{M}_t) = G(\mathbf{M}_t; \Sigma_{\mathbf{M}})$ where $G(\mathbf{M}_t; \Sigma_{\mathbf{M}})$ is a Gaussian model with a zero mean and the covariance matrix $\Sigma_{\mathbf{M}}$. The Gaussian prior model is used because of its simplicity and effectiveness of imposing smoothness constraints for 2D motions. Another component in Eqn. (1), the likelihood model $P(\mathbf{Z}_t|\mathbf{M}_t)$ measures the likelihood of motion parameters. In our method, the measurement of the model is a fusion of measurements of multiple devices, with more details provided in Section 4.2.

4.2 Integrated Model and Fusion of Individual Device Measurements

Due to the low image quality of fluoroscopy and cluttered backgrounds, independently tracking each device is prone to detection errors. To improve the accuracy and robustness of tracking, we combine all the devices into an "integrated IVUS model". We denote such a model as Γ_t, in which each component is represented by a curve (e.g., a segment of guiding catheter body and a wire) or a point (e.g., guiding catheter tip and IVUS transducer), as shown in Fig 4. By defining the integrated model, we can re-write the likelihood in the form of curve representations, as $P(\mathbf{Z}_t|\mathbf{M}_t) = P(\mathbf{Z}_t|\Gamma_t)$. Based on the integrated IVUS model, the measurement model is a combination of measurements of individual components. For simplicity, we assume that the measurement of individual component is independent of each other given the curve, i.e., $P(\mathbf{Z}_t|\Gamma_t^k, \Gamma_t) = P(\mathbf{Z}_t|\Gamma_t^k)$ in which Γ_t^k denotes a component in the integrated model. Therefore, we can further decompose the measurement model $P(\mathbf{Z}_t|\Gamma_t^k)$ as Eqn. (2):

$$P(\mathbf{Z}_t|\mathbf{M}_t) = P(\mathbf{Z}_t|\Gamma_t) = \sum_k P(\mathbf{Z}_t|\Gamma_t^k) P(\Gamma_t^k|\Gamma_t). \tag{2}$$

The component measurements $P(\mathbf{Z}_t|\Gamma_t^k)$ are from the four detectors, i.e. $P_{\text{IVUS\_trans}}$, $P_{\text{cath\_body}}$, $P_{\text{cath\_tip}}$ and P_{wire}. $P(\Gamma_t^k|\Gamma_t)$ defines the weight of individual component in the model, denoting the confidence of individual measurements. Such weights are set empirically in our method.

The integrated model to be tracked is initialized from the vessel detection at the angiogram phase. As shown in Fig. 4, the vessel detection identifies the vessel centerline, which is used to approximate the shape of wire. The guiding catheter body and tip is also initialized from the vessel detection. The IVUS transducer is initialized from the detection at the first frame in fluoroscopy. During tracking, the model is propagated from a previous frame, and updated from the motion parameter estimation. For computational efficiency, we apply a kernel-based multi-resolution method in the tracking [6] to achieve a speed of 1.0 frame per second in a computer with a Core 2 Duo 2.0GHz CPU.

4.3 Registration

The registration steps assume first the mapping of all tracked IVUS transducers to their corresponding positions along a vessel branch in angiography, and then the association of each IVUS image to the corresponding position in angiography. The first mapping between tracked IVUS transducers and selected vessel branch is based on the geodesic distance computed from the device tracking. The Euclidean distance between an IVUS transducer and a guiding catheter tip is directly computed from the tracking results, and then the geodesic distances can be inferred from the Euclidean distances based on the vessel shape. A constraint that the geodesic distance should change smoothly between successive frames is imposed to resolve potential ambiguity. Based on geodesic distances, each tracked IVUS transducer is mapped to a point on the vessel branch. After that, the association of IVUS images and angiography image is straightforward, with the use of the synchronized acquisition time of fluoroscopic and IVUS images.

5 Experiments

The method and system are evaluated on a set of clinical cases acquired from three different European hospitals. The acquisition follows the workflow described in Section 2. The evaluation set includes 65 cases. Each acquired case includes an angiogram sequence and a fluoroscopic pullback sequence (Artis Zee, Siemens AG, Forchheim, Germany), and corresponding IVUS images (Volcano Corp., San Diego, CA, U.S.A). The X-ray image size is 512 by 512, and the physical size of each pixel is between 0.2mm and 0.3mm. Each pullback takes a few minutes, usually containing 100 to 250 fluoroscopy frames. The data set includes a variety of clinical scenes, including low signal-to-noise ratio, different vessel branches and projection angles, and the existence of other surgical devices. Some examples are shown in Fig. 5.

To establish ground truth for quantitative evaluation, we manually annotate the positions of IVUS transducers and guiding catheter tips in the set as the

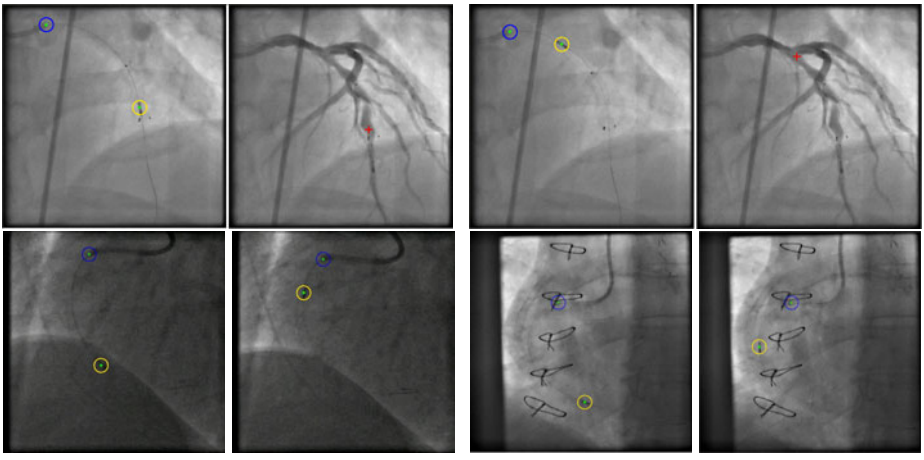


Fig. 5. Exemplar results: in each image of the first row, tracking results in the fluoroscopy are shown at the left, and the corresponding registration results in the angiogram are shown at the right; the second row shows more tracking results in fluoroscopy. IVUS images are not shown due to space limit.

ground truth. The tracking accuracy in fluoroscopy are quantitatively measured by the Euclidean distances between automatic tracking results and corresponding manual annotations. We further compute the geodesic distance error as the estimation of the registration error, because the final registration is based on the computed geodesic distance. For this purpose, we compute the geodesic distance from the manual annotations as the ground truth, and then compare this ground truth with the geodesic distance derived from corresponding automated tracking.

The quantitative validation of the system is performed with two experiments, i.e., Experiment 1 and Experiment 2. In the Experiment 1, the detectors are trained with and applied to all the cases. In the Experiment 2, the training and testing sets are separated to test how well our method can be generalized to unseen data. We train the detectors in 32 randomly selected cases, and validate the framework in the remaining 32 cases that are excluded from training. Our method is successful in 64 out of 65 cases, except for one case where the IVUS transducer is occluded by spines most time, thus achieving a 98.46% success rate. Some exemplar tracking results are shown in Fig. 5. Table 1 summarizes the error statistics from the 64 cases. The mm errors are converted from pixel errors based on the physical resolution of imaging detector, therefore quantifying normalized errors in a 2D projection space. The tracking error of IVUS transducers is small, with mean errors only around 0.35mm for both experiments. The error at the guiding catheter tips is slightly larger. The median error is 0.97mm for the Experiment 1, and 1.17mm for the Experiment 2. The overall small differences between the two experiments demonstrate the good generalization capability of our tracking method. The mean geodesic errors are below 1.20mm and 1.50mm

Table 1. Quantitative evaluation of the co-registration system

Validation	Experiment 1			Experiment 2		
Error	median	mean	std	median	mean	std
IVUS transducer (in pixels)	0.90	1.18	1.89	0.94	1.22	1.17
IVUS transducer (in mm)	0.26	0.34	0.56	0.28	0.35	0.32
Guiding catheter tip (in pixels)	3.34	5.43	6.42	4.10	6.72	7.43
Guiding catheter tip (in mm)	0.97	1.57	1.82	1.17	1.92	2.09
Geodesic distance (in pixels)	2.53	4.16	7.67	3.68	5.10	6.23
Geodesic distance (in mm)	0.73	1.19	2.14	1.03	1.41	1.51

from the two validation experiments respectively. The quantitative validations demonstrate that our method provides an accurate and robust co-registration of the angiography and IVUS images.

6 Conclusion

To our knowledge, this is the first reported system to provide automatic and robust tracking and registration of angiography and IVUS images. The experiments demonstrate its effectiveness in fusing the two imaging modalities.

We hope that this work will establish a useful reference for further research in this exciting field. The developed approach can be adapted to 3D when using biplane angiography and also be used in combination with other intravascular imaging modalities such as optical coherence tomography (OCT).

References

1. Bruining, N., de Winter, S., Serruys, P.W.: Intravascular ultrasound registration/integration with coronary angiography. Cardiology Clinics 27(3), 531–540 (2009)
2. Rieber, J., Prummer, S., Schmidt, M., Rittger, H.: A novel method for direct correlation of angiography and endovascular diagnostic tools in human coronary arteries in vivo. J. Am. Coll. Cardiol. 55, A217.E2065 ((2010)
3. Tu, S., Holm, N.R., Koning, G., Huang, Z., Reiber, J.H.C.: Fusion of 3D QCA and IVUS/OCT. Intl. J. Cardiovasc Imaging (2011)
4. Tu, Z.: Probabilistic boosting-tree: Learning discriminative models for classification, recognition, and clustering. In: ICCV, pp. 1589–1596 (2005)
5. Wahle, A., Prause, G., Birgelen, C.V., Erbel, R., Sonka, M.: Fusion of angiography and intravascular ultrasound in vivo: establishing the absolute 3-D frame orientation. IEEE Trans. on Biomedical Engineering (1999)
6. Wang, P., Chen, T., Zhu, Y., Zhou, S.K., Comaniciu, D.: Robust guidewire tracking in fluoroscopy. In: CVPR (2009)
7. Wang, P., Liao, W.-s., Chen, T., Zhou, S.K., Comaniciu, D.: Graph based interactive detection of curve structures in 2D fluoroscopy. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6363, pp. 269–277. Springer, Heidelberg (2010)

Nonmetallic Rigid–Flexible Outer Sheath with Pneumatic Shapelocking Mechanism and Double Curvature Structure

Siyang Zuo[1], Ken Masamune[1], Kenta Kuwana[1], Morimasa Tomikawa[2], Satoshi Ieiri[2], Takeshi Ohdaira[3], Makoto Hashizume[2,3], and Takeyoshi Dohi[1]

[1] Graduate School of Information Science and Technology, University of Tokyo, Japan
[2] Departments of Advanced Medicine and Innovative Technology, Kyushu University Hospital, Japan
[3] Departments of Advanced Medical Initiatives, Faculty of Medical Sciences, Kyushu University, Japan
{sasiyou,masa,k_kuwana}@atre.t.u-tokyo.ac.jp,
{mtomikaw,satoshi}@med.kyushu-u.ac.jp,
takeshiws.ohdaira@nifty.com, mhashi@dem.med.kyushu-u.ac.jp,
takdohi@i.u-tokyo.ac.jp

Abstract. Single port access (SPA) surgery is a laparoscopic procedure using only one transumbilical-placed port. Natural orifice transluminal endoscopic surgery (NOTES) offers the possibility of surgery without visible scars. To address the access and stability problems in SPA and NOTES, we developed a device called rigid–flexible outer sheath. This sheath can be switched between flexible and rigid modes by a novel pneumatic shapelocking mechanism, and it has a double curvature structure that enables it to flex in four directions at the distal end and three directions on the rigid–flexible shaft. The insertion part of the prototype is 300 mm long with a 20 mm outer diameter, and the part is equipped with four working channels. In vivo experiments using a swine show that the outer sheath has high potential for solving access and stability problems. We expect that the outer sheath will be useful for SPA and NOTES.

Keywords: Outer sheath, Pneumatic shapelocking mechanism, SPA, NOTES, MRI-compatible.

1 Introduction

Single port access (SPA) surgery is an advanced, minimally invasive surgical procedure using laparoscopic devices and instruments, and it is performed through a single entry point, which is typically the patient's navel. Because the entire procedure is conducted through the navel, it does not leave any visible scar. Recent advances in minimally invasive surgery have led to the natural orifice transluminal endoscopic surgery (NOTES) technique. Because no abdominal incisions are required, a major advantage of this method is the absence of associated abdominal wall complications. While the closure of the internal entry point for NOTES presents a significant

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 169–177, 2011.

challenge, patients largely prefer NOTES over standard laparoscopy because of the cosmetic benefits, except if the risk is significantly higher than that in the case of the laparoscopic approach [1].

Various endoscopic devices and instruments have been researched and developed to improve surgical dexterity and to reduce specific risks and difficulties associated with SPA. However, some problems remain unsolved. Firstly, SPA requires sufficient space below the abdominal wall for surgical procedures. Secondly, while SPA is useful when the affected area is on the anterior side of the body, it is difficult to approach the affected area if it is in a deep and narrow area or on the reverse side of organs. To overcome these complications, several research groups have developed flexible manipulators. Ikuta et al. developed a micromanipulator to approach inaccessible regions [2]. Other flexible manipulators have been developed using shape memory alloys [3] or a wire-driven mechanism [4]. However, flexible instruments cannot be easily inserted into the narrow spaces between tissues or organs and cannot be completely stabilized when approaching their targets. Although current endoscopes are highly advanced, numerous problems are encountered when they are used in NOTES. During NOTES, insertion of an instrument into an area of interest without causing injury to other tissues requires a surgical instrument with a flexible mode for insertion and a rigid mode for fixing the shaft in place. Many researchers have tried to develop systems using lockable sheaths or hybrid gaits involving locking and relaxing. Robert et al. applied this idea to medical applications [5]. A type of snake robot, called HARP [6], was also developed, and the TransPort (USGI Medical, San Capistrano, CA, USA) was designed for NOTES using ShapeLock technology [7]. However, these manipulators have some limitations. Firstly, because these manipulators use wire tension to lock the shape of the shaft, they often suffer from problems of wire breakage and thus cannot be used safely. Secondly, for locking the shape of the shaft, the surgeon must control at least three wires, which is a difficult maneuver. Thirdly, the mechanisms and structures of these manipulators are complicated and costly. Finally, there is no double curvature structure and MRI compatibility is difficult to achieve.

To solve these problems associated with access and stability in SPA and NOTES, we developed a novel outer sheath. The surgeon first inserts the outer sheath through the narrow gap between the safety areas. When the outer sheath approaches the target, the surgeon locks the shape and then inserts flexible instruments easily through the path created by the sheath. This paper reports the structure of the distal end, the pneumatic shapelocking mechanism, and the structure of the wire-control operating part. In addition, the paper presents a prototype of the outer sheath. We evaluated the performance of the outer sheath through *in vivo* experiments using a swine.

2 Outer Sheath Design

2.1 Structure of Integrated Pin-Joints at Bending Distal End

We designed a bending distal end using an integrated pin-joint structure. Frames 1, 2, 3, and 4 are joined together and driven by four wires that are 90° apart, which facilitates flexibility in four directions (**Fig. 1(A), (B)**).

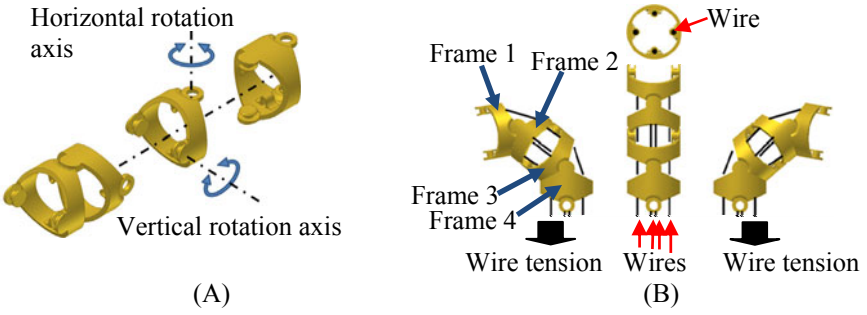


Fig. 1. Structure of bending distal end. (A) Image of units connected alternatively. (B) Mechanism of bending distal end.

The bending distal end consists of six aligned frames that mutually rotate 90° around their axes (**Fig. 4(C)**). This architecture facilitates a broad range of bending motions, resulting in ease of application during endoscopic surgery. In addition, the bending distal end is manufactured in an integrated manner, and therefore, assembly is not necessary. The integrated frame structures are produced by an Objet PolyJet-based 3D Printer (EDEN260V, Objet Geometries Ltd., Israel).

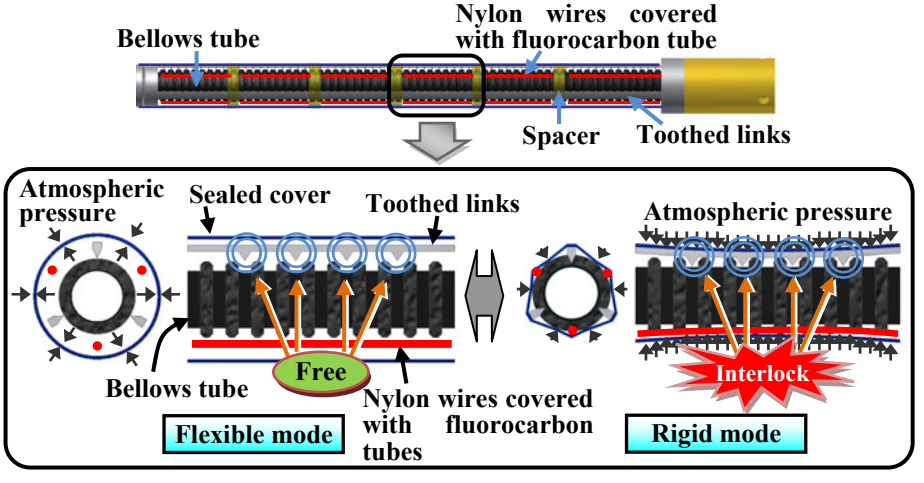


Fig. 2. Mechanism for switching between rigid and flexible modes. When there is no vacuum, the links and bellows tube can take any shape. When the pump creates a vacuum, the links move down and mesh with the ditch of the tube, and thus, the shape of the sheath is locked.

2.2 Mechanism of Pneumatic Shapelocking on Rigid–Flexible Shaft

With the mechanism described below, the outer sheath can alternate between flexible and rigid modes. The rigid–flexible shaft design consists of flexible toothed links, a bellows tube, a sealed cover, and nylon wires. In the flexible mode, the internal and

atmospheric pressures equalize. Therefore, the sealed cover does not shrink, and the toothed links disengage from the bellows tube. In the rigid mode, the sealed space is evacuated by discharging the internal air, and the atmospheric pressure presses the toothed links into the bellows tube. Thus, the shape of the outer sheath is locked (**Fig. 2**). The bellows tube and toothed link mechanism can be easily locked as well as relaxed, providing a smooth transition between flexible and rigid modes. Furthermore, three nylon wires covered with non-shrinkable fluorocarbon tubes pass through the clearance space along the bellows tube to bend the rigid–flexible shaft of the outer sheath.

2.3 Structure of Chain-Gear on Wire-Control Operating Part

The wire-control operating part for the bending distal end has a chain-gear structure (**Fig. 3**). A surgeon can rotate the two knobs to easily control the bending angle of the distal end. Bending angles of ±120° in the vertical direction and ±90° in the horizontal direction can be achieved by Knob 1 (**Fig. 3**) from -60° to +60° and Knob 2 (**Fig. 3**) from -70° to +70°, respectively. To realize MRI compatibility, the materials of the operating part are made of plastic.

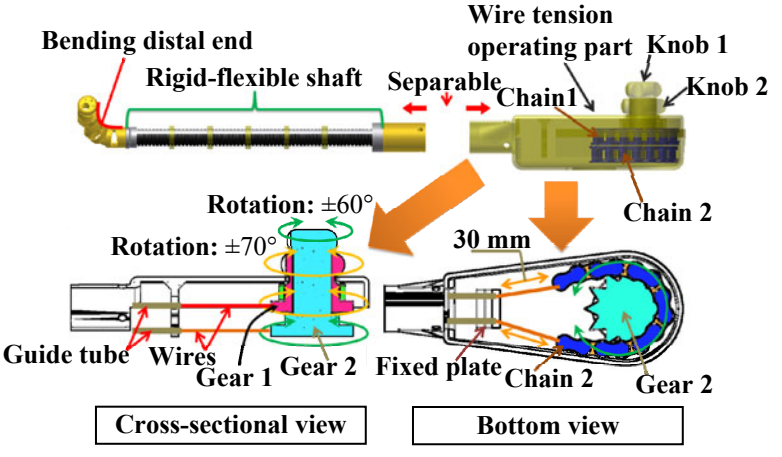


Fig. 3. Wire-control operating part for the bending distal end. Four nylon wires are connected by the chains. The bending angle of the distal end is controlled by two sets of chain–gear structures.

2.4 Prototype

We built a prototype of the outer sheath (**Fig. 4(A), (B)**). The prototype has a length of 300 mm, maximum outer diameter of 20 mm. The length of the bending distal end is 75 mm. In addition, our model was equipped with one 7-mm, one 2-mm, and two 1.35-mm working channels. The flexible instruments can be inserted from the 7-mm and 2-mm channels, and the two 1.35-mm channels are used for water jet and suction. The rigid–flexible shaft consists of three long, flexible toothed links, a bellows tube,

and a polyethylene cover. The toothed links are 120° apart, which allows the shaft to be locked in any direction. The bending distal end consists of six frames and four nylon wires. For each frame, the rotating angle in the vertical and horizontal direction is ± 40° and ±45°, respectively, because of which the bending distal end can achieve a curvature of ±120° and ±90° in the vertical and horizontal directions, respectively (**Fig. 4(C)**). The rigid–flexible shaft also has active bending capability because of three nylon wires that are 120° apart, and the maximum bending angle achieved by pulling the wires is 90°. The wires pass through three holes in the operation part and are manually controlled. The system of the prototype is shown in **Fig. 4(D)**. The outer sheath is separated from the vacuum controller and the vacuum source to be cleaned and sterilized. All parts of the prototype are made of plastic, and this ensures excellent MRI compatibility.

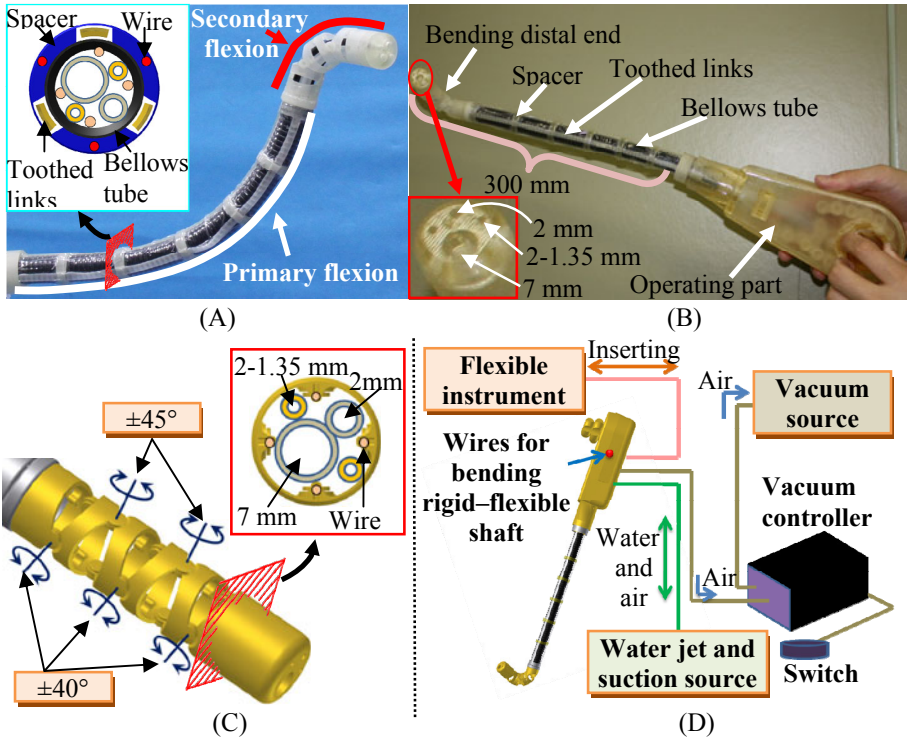


Fig. 4. Prototype of the outer sheath. (A) The outer sheath can curve like a snake and hold its shape. (B) Image of prototype. (C) Bending distal end. (D) System configuration.

2.5 Mechanical Characteristics of Prototype

We first examined the bending characteristics of the distal end by pulling four wires. A bending angle of ±90° in the horizontal direction was achieved through a wire displacement of 25 mm and a wire tension of 8.49 ± 0.23 N (n = 3). A bending angle

of ±120° in the vertical direction was achieved through a wire displacement of 30 mm and a wire tension of 11.03 ± 0.26 N (n = 3). We next tested the bending characteristics of the rigid–flexible shaft by pulling the three wires. A maximum bending angle of 90° was achieved through a wire displacement of 70 mm and a wire tension of 12 ± 0.69 N (n = 3). Finally, we tested the maximum holding power of the sheath in maintaining its shape using only the pneumatic shapelocking mechanism in the rigid mode. The limited holding power was determined by visually observing the point at which deformation of the outer sheath began. The holding powers were measured in two conditions: the straight condition (0°) (**Fig. 5(A)**) and curved condition (90°) (**Fig. 5(B)**). In both cases, the load direction was the downward vertical direction. The maximum torque for which deformation of the shape did not occur was 29.55 N•cm in case (A) and 21.75 N•cm in case (B). In the flexible mode, the shape of the rigid–flexible shaft can be changed under a torque of 12 N•cm in the straight condition (0°).

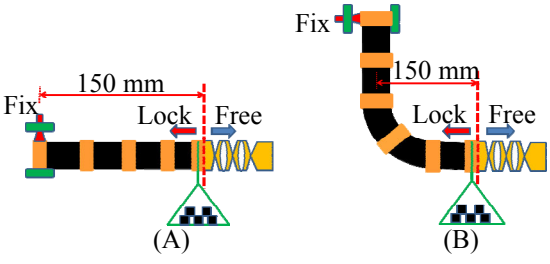


Fig. 5 Measurement of holding power in the rigid mode. (A) Load in the straight condition (0°). (B) Load in the curved condition (90°).

3 *In Vivo* Experiment

In the *in vivo* experiment, we tested the efficacy of our prototype in the abdominal cavity of a swine (male, 48.5 kg). The surgery was performed through two access ports. We inserted a laparoscope into the port in the lower abdominal region to observe and confirm the position and shape of the outer sheath, and we inserted the outer sheath through the center of the abdominal region. The incision part during insertion of the outer sheath was protected with a LAP DISK (Hakko Medical Inc., Japan). The instruments inserted into the outer sheath were a 2.8-mm fiberscope (Sumita Optical Glass, Inc., Japan) and a 1.75-mm biopsy forceps (BF1812SF, Fujinon Toshiba ES Systems Co. Ltd., Japan). The outer sheath was connected to a vacuum pump (DTC-41, ULVAC KIKO Inc., Japan) and a vacuum controller to alternate between flexible and rigid modes by pushing a button. The sheath was also connected to a roller pump (RP-2100, Tokyo Rikakikai Co., Ltd., Japan) to jet water. Suction was applied through a vacuum supply port in the operating room.

First, we inserted the outer sheath into the gap between the liver and diaphragm to observe the posterior side of the liver with a 2.8-mm fiberscope (**Fig. 6(A)**). Second, we inserted the outer sheath into the abovementioned gap to locate the bladder, and we clamped the bladder surface using a biopsy forceps (**Fig. 6(B)**). Third, we sprayed

water from the 1.35-mm channel to clean the lens of the fiberscope and the internal organ, and then, we drew the water out through the second 1.35-mm channel (**Fig. 6(C)**). Fourth, we inserted the outer sheath into the gap among the spleen, liver, and stomach. These tasks were performed without the use of forceps (**Fig. 6(D)**). Finally, we tried to approach the colon from the posterior side (**Fig. 6(E)**). Furthermore, this prototype can be used in the rigid mode as a rigid instrument similar to conventional laparoscopic devices (**Fig. 6(F)**).

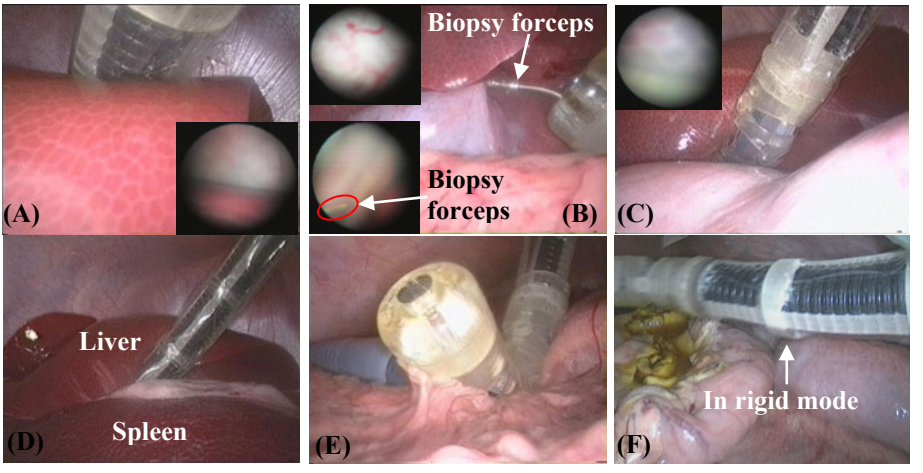


Fig. 6. Laparoscopic broad views along with fiberscope local views of the abdominal cavity of a swine. (A) Image of the outer sheath entering the gap between the liver and diaphragm. The fiberscope shows the border between the organs. (B) Image showing the clamping of liver lobes by biopsy forceps inserted through the outer sheath. The fiberscope shows the surface of the liver lobes and biopsy forceps. (C) Image showing the outer sheath during underwater observation. The fiberscope shows the underwater image. (D) The outer sheath entering the gap among the spleen, liver, and stomach. (E) The colon approached with a widely curved distal end. (F) The prototype can perform spatial migration like rigid instruments in rigid mode.

4 Discussion and Conclusion

We confirmed that the bending distal end of the outer sheath enabled switching between ±120° and ±90° curvatures in the vertical and horizontal directions, respectively. The outer sheath facilitates insertion of flexible instruments into gaps between organs; therefore, even a torque of 29.55 N•cm for a straight shape and 21.75 N•cm for 90° curved shapes may be sufficient against external pressure from organs and internal pressure from inserted instruments through the channels. This belief was validated through *in vivo* experiments. The locking mechanism guards against unintentional activation created by organs pressing against the shaft. Because the outer sheath employs vacuum as the locking mechanism at low pressures of less than 1 kPa, the tissues near the outer sheath are protected from damage even in the event of air leakage. The minimum bend radius of the rigid–flexible shaft is constrained by the

minimum bend radius (50 mm) of the bellows tube. This radius corresponds to a slightly more curved path for maneuvering around small organs. However, the radius is sufficient for approach between abdominal walls and tissues or for maneuvers around large organs such as the liver.

In vivo evaluations showed that the outer sheath has strong potential for solving access problems and stability issues. The double curvature and locking capability allow the outer sheath to be positioned at the target field and to lock the shape of the rigid–flexible shaft. Thus, the surgeon's hands are free. These features, combined with the endoscope's four-directional flexion, enable more complex manipulation. The outer sheath has a separate channel for water jet and suction. A 7-mm channel is designed for a 6-mm flexible endoscope. The low resolution of the 2.8-mm fiberscope was not adequate for observing local fields, and the 2-mm channel is not sufficiently large for standard devices such as endoscopic clips or more robust instruments. Therefore, a high-resolution scope and larger channel for robust devices are necessary for more complicated procedures. Because the locking principle of the outer sheath is based on vacuum and all parts of this device are made of plastic, the outer sheath should have little influence on MRI, which enables MRI-guided surgery [8]. Surgeons recommended that the device be used for natural orifice NOTES procedures, especially for transrectal and transvaginal NOTES, because the approach is suitable for short, gentle curves. The performance of the outer sheath with improvements in channel size and triangulation will be evaluated through future *in vivo* experiments on NOTES.

To secure a stable approach path for SPA and NOTES, we developed a nonmetallic rigid–flexible outer sheath that has four working channels. This sheath can flex in four directions at the bending distal end and three directions on the rigid–flexible shaft. Through *in vivo* experiments using a swine, we determined that flexible instruments can be inserted into deep areas in the abdominal cavity and locked into position, enabling independent flexion at the distal end. Thus, the design of the outer sheath meets the requirements for SPA and NOTES. The results of our study show that the sheath has potential for application in SPA and NOTES.

References

1. Varadarajulu, S., Tamhane, A., Drelichman, E.: Patient perception of natural orifice transluminal endoscopic surgery as a technique for cholecystectomy. J. Gastrointest. Endosc. 67, 854–860 (2008)
2. Ikuta, K., Sasaki, K., Yamamoto, K., Shimada, T.: Remote Microsurgery System for Deep and Narrow Space - Development of New Surgical Procedure and Micro-robotic Tool. In: Dohi, T., Kikinis, R. (eds.) MICCAI 2002. LNCS, vol. 2488, pp. 163–172. Springer, Heidelberg (2002)
3. Nakamura, Y., Matsui, A., Saito, T.: Shape Memory-Alloy Active Forceps for Laparoscopic Surgery. In: Proc. 1995 IEEE International Conference on Robot and Automation, vol. 3, pp. 2320–2327 (1995)
4. Simaan, N., Taylor, R., Flint, P.: High Dexterity Snake-Like Robotic Slaves for Minimally Invasive Telesurgery of the Upper Airway. In: Barillot, C., Haynor, D.R., Hellier, P. (eds.) MICCAI 2004. LNCS, vol. 3217, pp. 17–24. Springer, Heidelberg (2004)

5. Sturges, R.H., Laowattana Jr., S.: A flexible, tendon-controlled device for endoscopy. In: Proc. 1991 IEEE International Conference on Robotics and Automation, vol. 3, pp. 2582–2591 (1991)
6. Amir, D., Howie, C., Alon, W., Takeyoshi, O., Marco, A.Z.: Percutaneous intrapericardial interventions using a highly articulated robotic probe. In: The First IEEE/RAS-EMBS International Conference on Biomedical Robotics and Biomechatronics, pp. 7–12 (2006)
7. Raju, G.S., Rex, D.K., Kozarek, R.A., Ahmed, I., Brining, D., Pasricha, P.J.: A novel shape-locking guide for prevention of sigmoid looping during colonoscopy. J. Gastrointest. Endosc. 59, 416–419 (2004)
8. Zuo, S., Yamanaka, N., Sato, I., Masamune, K., Liao, H., Matsumiya, K., Dohi, T.: MRI-Compatible Rigid and Flexible Outer Sheath Device with Pneumatic Locking Mechanism for Minimally Invasive Surgery. In: Dohi, T., Sakuma, I., Liao, H. (eds.) MIAR 2008. LNCS, vol. 5128, pp. 210–219. Springer, Heidelberg (2008)

3D Stent Recovery from One X-Ray Projection

Stefanie Demirci[1,⋆], Ali Bigdelou[1,⋆], Lejing Wang[1], Christian Wachinger[1], Maximilian Baust[1], Radhika Tibrewal[1], Reza Ghotbi[2], Hans-Henning Eckstein[3], and Nassir Navab[1]

[1] Computer Aided Medical Procedures, Technische Universität München, Germany
{demirci,bigdelou,wangle,wachinge,baust,rtibrewal,navab}@cs.tum.edu
[2] Vascular Surgery, Klinikum München-Pasing, Germany
reza.ghotbi@krankenhaus-pasing.de
[3] Department of Vascular Surgery, Klinikum rechts der Isar, Germany
h.h.eckstein@lrz.tum.de[⋆⋆]

Abstract. In the current clinical workflow of endovascular abdominal aortic repairs (EVAR) a stent graft is inserted into the aneurysmatic aorta under 2D angiographic imaging. Due to the missing depth information in the X-ray visualization, it is highly difficult in particular for junior physicians to place the stent graft in the preoperatively defined position within the aorta. Therefore, advanced 3D visualization of stent grafts is highly required. In this paper, we present a novel algorithm to automatically match a 3D model of the stent graft to an intraoperative 2D image showing the device. By automatic preprocessing and a global-to-local registration approach, we are able to abandon user interaction and still meet the desired robustness. The complexity of our registration scheme is reduced by a semi-simultaneous optimization strategy incorporating constraints that correspond to the geometric model of the stent graft. Via experiments on synthetic, phantom, and real interventional data, we are able to show that the presented method matches the stent graft model to the 2D image data with good accuracy.

1 Introduction

Due to its minimally invasive nature, endovascular aortic repair (EVAR) has replaced the conventional open surgery as preventive treatment procedure for abdominal aortic aneurysms (AAA). In contrast to a complete opening of the patient's abdomen, a stent graft is inserted through one femoral artery into the aneurysmatic aorta excluding the aneurysm sack from the circulation and reducing the pressure on the aortic wall. Once completely unfolded, misplacements of the stent graft may lead to life-threatening emergency surgeries if vitally important organs are cut-off from blood supply.

⋆ Joint first authors.

⋆⋆ We are thankful to Simon Weidert (Klinikum Innenstadt, Ludwig-Maximilian-Universität München, Germany) and Asbjorn Odegard (St. Olavs Hospital Trondheim, Norway) for providing image data.

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 178–185, 2011.

Fig. 1. Tubular aortic stent graft

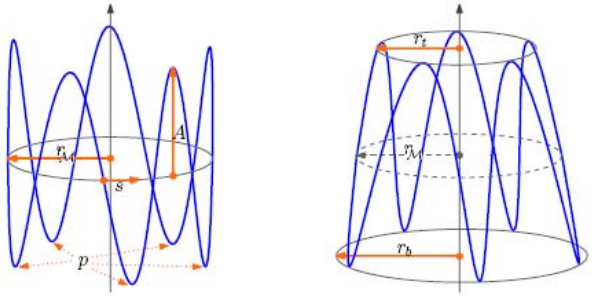


Fig. 2. Stent segment model with $r^t = r^b$ (left) and $r^t < r^b$ (right)

During the intervention, the catheter and stent position is only visualized in 2D X-ray views. This makes it highly difficult for physicians to place the stent graft in the predefined position. A virtual visualization of the unfolding stent graft within the CTA volume would provide the physician a 3D view of the current situation. Such solutions would decrease both the need for extensive use of contrast agent and the radiation dose. Whereas 2D-3D registration systems for endovascular interventions are already available [6,8], a visualization of the stent graft in a 3D patient scan still requires a method for stent graft detection in 2D and correct backprojection into 3D.

Existing publications on stent graft segmentation in 3D volumes [7,4] all employ intensity based thresholding. There is, to the best of our knowledge, no shape-based approach for extraction of aortic stent grafts in 3D or 2D.

In this paper, we present a novel algorithm to match a 3D model of the stent graft to an intraoperative 2D image showing the device.

2 Method

2.1 Stent Model

We define the stent segment model to be the curve $\mathcal{M}_i(\mathbf{x}) = (a_i(\mathbf{x}), b_i(\mathbf{x}), c_i(\mathbf{x}))$ consisting of the set of parametric equations

$$a_i(\mathbf{x}) = r_{\mathcal{M},i}(\mathbf{x})cos(\mathbf{x}) \tag{1}$$
$$b_i(\mathbf{x}) = r_{\mathcal{M},i}(\mathbf{x})sin(\mathbf{x}) \tag{2}$$
$$c_i(\mathbf{x}) = A_i sin(p_i\mathbf{x} + s_i) \tag{3}$$

with amplitude A_i specifying the height of the segment, period p_i equal to the number of peaks, and phase shift s_i merely shifting the starting point. The radius of the stent segment model is calculated by

$$r_{\mathcal{M},i}(\mathbf{x}) = \left(1 - \frac{\sin(p_i\mathbf{x}+1)}{2}\right) r_i^t + \frac{\sin(p_i\mathbf{x})+1}{2} r_i^b \tag{4}$$

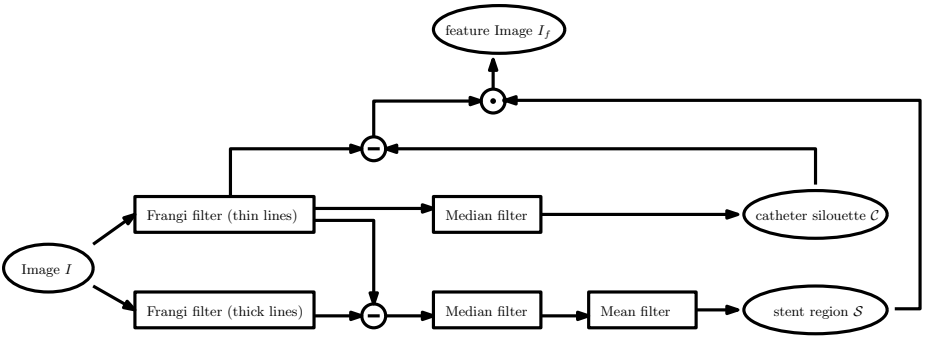
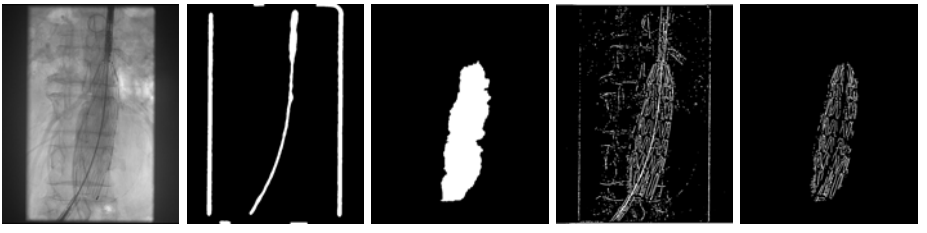


Fig. 3. Preprocessing scheme for an automatic computation of the feature image I_f



(a) I (b) $\mathcal{C}$ (c) $\mathcal{S}$ (d) I_{wires} (e) I_f

Fig. 4. Visualization of automatic feature extraction procedure

with r_i^t and r_i^b representing the upper and lower radius as visualized in Fig. 2. The entire stent graft is defined as the set of l segments $\mathcal{M} = \{\mathcal{M}_1, \ldots, \mathcal{M}_l\}$.

2.2 Automatic Feature Extraction

For an improved performance, we first apply a preprocessing procedure on image I (see Fig. 4(a)) in order to extract a region of interest and highlight the wire structure. Additional to the wires of the stent graft, the guide wire which is used as navigation support to the physician is also visible in the interventional X-ray images. In order for our matching algorithm to not get stuck in these outliers, we need to eliminate these pixels.

Fig. 3 displays our preprocessing scheme. For obtaining the catheter silhouette image $\mathcal{C}$ (Fig. 4(b)), we employ the Frangi filter [2] for scales $5-6$ followed by a median filtering for noise removal in order to capture the catheter pixels.

The stent region $\mathcal{S}$ (Fig. 4(c)) is extracted similarly. First, we subtract thick curvilinear structures from thin curvilinear structures (Frangi filter for scale 2) for only highlighting the stent wires in I_{wires} (Fig. 4(d)). Subsequent employment of a median filter for noise removal and mean filter for dominant region extraction leads to the desired image region that contains the stent graft. Instead of the Frangi filter, any filter can be used that highlights curvilinear structures of selected sizes.

In the end, we subtract the catheter pixels $\mathcal{C}$ from image I_{wires} and mask the result with $\mathcal{S}$. The resulting feature image I_f is displayed in Fig. 4(e). In order to provide a distance measure for the following registration procedure, we calculate distance map D_{I_f} on feature image I_f.

2.3 Stent-Model-to-Image Registration

A projection is necessary to map the stent model $\mathcal{M}$ to the 2D image coordinate system of I. Similar to 2D-3D image registration, the projection transformation $\mathbf{P} = K[R|t]$ consists of the 6-DOF extrinsic parameters $[R|t]$ for rotation and translation of the 3D volume and the 4-DOF intrinsic camera parameter K of the *pinhole camera model* [3]. For the following considerations, we assume the camera matrix K to be given by the interventional angio system.

Using all transformation and model parameters together as

$$\mathbf{p} = \{\mathbf{p}_1, \ldots, \mathbf{p}_l\}, \qquad \mathbf{p_i} = \{R_i, t_i, r_i^t, r_i^b, A_i\} \tag{5}$$

the registration problem for the entire stent model is given by

$$\hat{\mathbf{p}} = \arg\min_{\mathbf{p}} \sum_{i=1}^{l} \sum_{x \in \mathcal{M}_i} D_{I_f}(T_{\mathbf{p}_i}(x)) \tag{6}$$

where $T_{\mathbf{p}_i}(x)$ is a projection of point x of the 3D segment model $\mathcal{M}_i$ using parameters $\mathbf{p}_i$. This equals a simultaneous registration of all stent segment models introducing a parameter space of dimension $l \times 9$. Considering that conventional abdominal aortic stent grafts consist of several segments, the cost for optimization increases rapidly. Another drawback is the presence of many local minima in the costfunction plot corresponding to each one of all displayed stent segments.

In order to reduce the complexity in our registration procedure, we employ a semi-simultaneous optimization framework based on the idea of Sidorov et al. [9]. Instead of optimizing all parameters for all segments at once, we optimize the parameters of one segment for a certain number of iterations and then move to the next randomly chosen segment. By applying this strategy, we implicitly make use of the tubular appearance of the stent graft.It requires however a fairly good initial guess of the position of the stent in the image which we try to achieve by separating the pose estimation process into a global and a local step.

Global registration. Here, we solve for the overall orientation of all segments in order to be very close for the local calculations. The global pose of the entire stent graft model $\mathcal{M}$ is defined by the global parameters $\mathbf{p}_{\text{global}} = \{K, R_{global}, t_{global}\}$. Assuming that K is given, a good initial guess for the remaining global parameters is achieved by computing a *principal component analysis* and subsequent region operations on the stent region $\mathcal{S}$.

Accordingly, let now $\mathbf{p}_i = \{\mathbf{v}_i, \mathbf{t}_i, r_i^t, r_i^b, A_i\}$ define the set of remaining parameters for each single segment i, where $\mathbf{v}_i = [\alpha_i, \beta_i, \gamma_i]^T$ represents the vector containing the three rotation angles that form rotation matrix $\mathbf{R}_i$. Setting

parameter vectors $\mathbf{p}_i$ to initial values with no rotation and translation and predefined approximated values for A and $r^t = r^b$, we obtain an estimation of the position of each stent segment in the interventional image.

Local registration. In this step, values for each $\mathbf{p}_i$ are optimized. Similar to Equ. (6), we define our costfunction for each segment i $(i = 1, \ldots, l)$ as

$$E(\mathbf{p}_i) = \sum_{x \in \mathcal{M}_i} D_{I_f}(T_{\mathbf{p}_{global} \circ \mathbf{p}_i}(x)) + \frac{\omega}{\lambda} \langle \phi(\mathbf{p}_i), \phi(\mathbf{p}_{i+1}) \rangle \tag{7}$$

with

$$\phi(\mathbf{p}_i) = \left\| \begin{array}{c} |\mathbf{t}_i - \mathbf{t}_{i-1}| - \mathbf{t}_i^{\Delta} \\ |\mathbf{v}_i - \mathbf{v}_{i-1}| - \mathbf{v}_i^{\Delta} \\ |r_i^t - r_{i-1}^b| - r_i^{\Delta} \end{array} \right\| \tag{8}$$

constraining the change of translation, rotation and radius in between neighboring segments naturally defined by the graft material fixing the stent wires (Fig. 1) that is not stretchable. Therefore respective parameters depend on the predefined distance d_i between the segments:

$$\mathbf{t}_i^{\Delta} = \begin{bmatrix} \frac{d_i}{2} \\ \frac{d_i}{2} \\ \frac{A_i}{2} + d_i \end{bmatrix} \qquad \mathbf{v}_i^{\Delta} = \begin{bmatrix} \frac{d_i}{\sqrt{1+d_i}} \\ \frac{d_i}{\sqrt{1+d_i}} \\ \frac{s_i}{2} \end{bmatrix} \qquad r_i^{\Delta} = |r_i^t - r_i^b|$$

$\omega >> 0$ represents a weighting factor for the penalization term. We optimize Equ. 7 within a deterministic annealing scheme with temperature λ being the distance from the current projected segment to the nearest image feature in I_f. As the distance and hence the temperature decrease, the penalization term is taken more and more into account. Using this approach our algorithm is also able to capture highly curved stent grafts.

3 Results and Discussion

Although our method can be applied to interventional X-ray images without knowing the corresponding matrix $\mathbf{P}$, a thorough validation is only possible in 3D requiring a correct calibration of the C-arm system. A possible strategy is to acquire projection images of the stent graft from two or more different views, apply our proposed method to each of the images, and compare the resulting 3D models. Therefore it is crucial that either all images are taken simultaneously or no changes to the stent graft have been made in between the different acquisitions. It is very difficult to find interventional images that fulfill these requirements as bi-planar systems are not considered state-of-the-art imaging for endovascular interventions and physicians only rarely acquire two or more images where the stent graft is in the exact same opening stage. Another approach would be to compare the resulting 3D stent model to a segmentation of the stent in a 3D volume. This is not possible for interventional cases, as only

Table 1. Synthetic experiments: RMS errors for amplitude, top and bottom radii averaged over total number of included segments and target registration errors over all landmarks

	A (mm)	r^t (mm)	r^b (mm)	# landmarks	mean/median/max TRE (mm)
$\mathbf{R}_3$	0.00	0.00	0.00	12	6.29/3.86/8.65
$\mathbf{R}_4$	0.00	0.00	0.00	16	6.54/4.01/7.96
$\mathbf{R}_5$	0.00	0.04	0.02	20	7.02/5.36/10.50
$\mathbf{R}_6$	0.01	0.04	0.04	24	7.39/5.81/9.91

pre- and (maybe) postoperative patient volumes are available that either contain no stent graft at all (preoperative) or a fully opened stent graft (postoperative). Due to these difficulties, we decided to quantitatively validate the accuracy of our proposed algorithm on realistic synthetic and phantom data and use in-vivo data to show the interventional impact of our method.

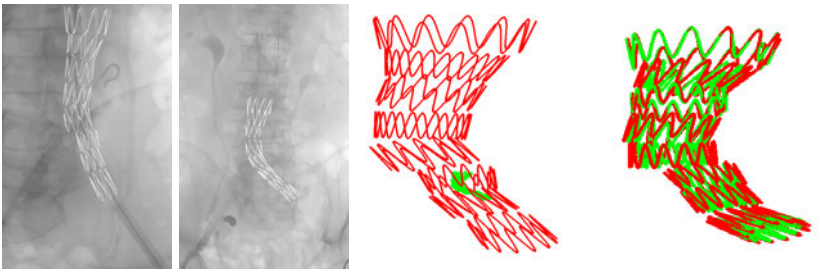
With this paper we intend to present a proof-of-concept for a highly novel approach to enhance intraoperative visualization for EVAR. We would like to emphasize that the proposed method is of linear complexity $\mathcal{O}(n)$ with n being the number of pixels included in one stent segment. Depending on the image resolution, n lies in between 100 and 500. Although all experiments were performed using unoptimized MATLAB code, with an efficient implementation, it is possible to apply the procedure within an interventional setting.

For creating the synthetic data, we digitally produced binary volumes of size $512 \times 512 \times 512$ showing random stent graft models consisting of 3-6 segments, namely $\mathbf{R}_3, \mathbf{R}_4, \mathbf{R}_5, \mathbf{R}_6$. By applying realistic deformations to each of the segments and calculating random projections of the volumes, we obtain 2D images of the stent graft as well as the corresponding ground truth parameter selection.We estimated the average noise level of 5 interventional fluoroscopies showing an unfolded stent graft and added it to the projection images (mean 0.59 and variance 0.98). The registration accuracy measured in *target registration error* (TRE) in varying number of distributed landamrks, is listed in Table 1. As expected, the errors values are very small and show the overall impact of our method. The sudden change in error values for stent models with 5 and 6 segments are due to the fact that we only employed different top and bottom radii for $\mathbf{R}_5, \mathbf{R}_6$.

The phantom design is composed of the upper part of a swine's leg in which we inserted a halfly unfolded stent graft parallel to the bone. For guide wire simulation, we integrated a small wire inside the stent graft. In order to prevent deformation changes, we placed the entire phantom in a paper box and filled all empty space with insulating foam. 10 radio-opaque markers were sticked to the outside and inside of the box in a predefined pattern.After acquiring a 3D CT volume, we took X-ray projection images with a mobile interventional C-arm from 3 different angles. Details of the image sizes and transformations are given in Table 2. In absence of an available fixed angio system, we had to calibrate the C-arm by employing the *normalized Direct Linear Transform (DLT)* algorithm [3] on the corresponding marker positions. Once $\mathbf{P}$ matrices were estimated, our algorithm was applied to all 3 projection images. We compared the recovered

Table 2. Details for phantom projection images and the Hausdorff distance of recovered 3D stent model to real stent graft extracted from CT

	size (pixel)	1st angle (deg)	2nd angle (deg)	Hausdorff (mm)
A	478 × 422	18.5	20.0	20.2
B	478 × 422	0.5	15.0	25.4
C	478 × 422	108.5	−4.5	15.1



(a) (b) (c) (d)

Fig. 5. Real experiment: (a)-(b) detection results (white) overlayed onto interventional images acquired from two different views, (c) 3D recovery of both stent shapes (green,red) and (d) after applying a common scale to all green segments

3D models to the real stent extracted from CT using a combination of manual outlining and region growing. In order to allow for quantitative evaluation of our results, we backprojected the resulting digital stent models into CT space and created a binary volume of the exact same extent and voxel size as the phantom CT volume. Values for the Hausdorff distance measuring the distance between both shapes for each of the 3 datasets, are given in Table 2.

For the real experiments, our medical partners kindly provided a set of real interventional fluoroscopy image data (acquired by Siemens AXIOM Artis dTA angiography suite) and corresponding preoperative CTA scans. Provided that the C-arm system is correctly calibrated, the amplitude and top and bottom radii of two corresponding segments acquired from different views must be equal up to a common scaling factor. As the angiography system used by our clinical partners is not biplanar, deformation changes need to be taken care of by the experimental setup. In the lower abdominal part of the human body, the aorta and iliac arteries are not exposed to breathing or other organ specific motion and the pulsatile motion originating from the blood pressure is neglectable [5]. The deformation that is induced by the stent graft itself, however, is very significant [1], but can be eliminated by acquiring both images in the same opening stage of the stent graft. Visual results are given in Fig. 5. For the combined plot of both models (Fig. 5(d)), we applied a common scale of 7.91 to each segment's amplitude and top and bottom radii of the stent model shown in red color. The scale value represents the mean ratio of final values for amplitude and top and bottom radii for each of the 8 segments and yields a root mean square error in model point positions of 2.1 mm.

4 Conclusion

In this paper, we presented a novel algorithm to match a 3D model of the stent graft to an intraoperative 2D image showing the device. It uses automatic preprocessing and a global-to-local approach to abandon any user interaction and still meet the required robustness. Including a semi-simultaneous optimization strategy and constraining the inter-stent-segment relations to correspond to the material property of the stent graft, the complexity of the optimization space could be reduced. The results of performed experiments on synthetic as well as real interventional data show the practical potential of our proposed method.

References

1. Demirci, S., Manstad-Hulaas, F., Navab, N.: Quantification of aortic deformation after EVAR. In: Miga, M.I., Wong, K.H. (eds.) Medical Imaging: Visualization, Image-Guided Procedures, and Modeling. Proceedings of SPIE. SPIE, vol. 7261, p. 72611U (2009)
2. 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)
3. Hartley, R., Zisserman, A.: Multiple View Geometry, 2nd edn. University Press, Cambridge (2003)
4. Klein, A., Renema, W., Schultze Kool, L., Slump, C.: Initial steps towards automatic segmentation of the wire frame of stent grafts in ct data. In: 4th Annual Symposium of the IEEE-EMBS Benelux Chapter, pp. 116–119. IEEE-EMBS Benelux Chapter, Enschede (2009)
5. Klein, A., Oostveen, L.J., Greuter, M.J.W., Hoogeveen, Y., Schultze Kool, L.J., Slump, C.H., Renema, W.K.J.: Detectability of motions in aaa with ecg-gated cta: a quantitative study. Med. Phys. 36(10), 4616–4624 (2009)
6. Liao, R., Tan, Y., Sundar, H., Pfister, M., Kamen, A.: An efficient graph-based deformable 2D/3D registration algorithm with applications for abdominal aortic aneurysm interventions. In: Liao, H., Edwards, P.J., Pan, X., Fan, Y., Yang, G.-Z. (eds.) MIAR 2010. LNCS, vol. 6326, pp. 561–570. Springer, Heidelberg (2010)
7. Mattes, J., Steingruber, I., Netzer, M., Fritscher, K., Kopf, H., Jaschke, W., Schubert, R.: Quantification of the migration and deformation of abdominal aortic aneurysm stent grafts. In: Reinhardt, J.M., Pluim, J.P.W. (eds.) Medical Imaging: Image Processing. Proceedings of SPIE, vol. 6144, p. 61440V. SPIE (2006)
8. Raheem, A., Carrell, T., Modarai, B., Penney, G.: Non-rigid 2D-3D image registration for use in endovascular repair of abdominal aortic aneurysms. In: Bhalerao, A.H., Rajpoot, N.M. (eds.) Proceedings of Medical Image Understanding and Analysis, pp. 153–158. University of Warwick (2010)
9. Sidorov, K., Richmond, S., Marshall, D.: An efficient stochastic approach to group-wise non-rigid image registration. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 2208–2213. IEEE Press, Los Alamitos (2009)

Augmented Virtuality for Arthroscopic Knee Surgery

John M. Li[1], Davide D. Bardana[2], and A. James Stewart[1,2]

[1] School of Computing, Queen's University
[2] Human Mobility Research Centre, Kingston General Hospital

Abstract. This paper describes a computer system to visualize the location and alignment of an arthroscope using augmented virtuality. A 3D computer model of the patient's joint (from CT) is shown, along with a model of the tracked arthroscopic probe and the projection of the camera image onto the virtual joint. A user study, using plastic bones instead of live patients, was made to determine the effectiveness of this navigated display; the study showed that the navigated display improves target localization in novice residents.

1 Introduction

Arthroscopic knee surgery is a minimally invasive procedure in which the surgeon navigates a surgical tool using camera images displayed on a screen above the patient. A small incision allows the arthroscope to provide a view of the surgical site while inducing less trauma than comparable open surgery.

However, navigating within the joint is challenging because the camera image on the overhead display has an unintuitive relationship with the arthroscope in the surgeon's hand, making hand/eye coordination very difficult. Surgeons must make a mental coordinate transformation to become correctly oriented within the patient.

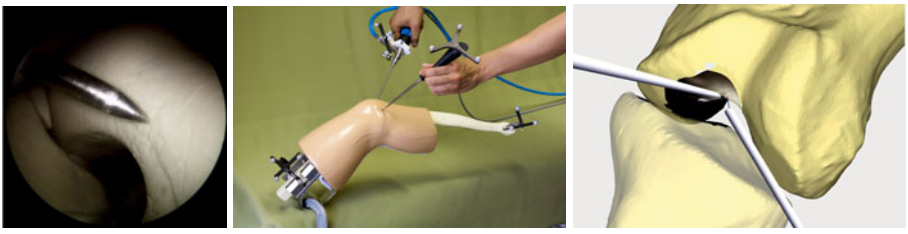
This paper describes a navigation system to visualize the arthroscope and probe in relation to a virtual model of the patient's joint (Figure 1). The navigation system also shows the camera image projected onto the virtual joint to make explicit the relationship between the camera image and the patient's anatomy.

This approach is referred to as "augmented virtuality" (AV), where a virtual view is augmented with real world camera images. In the navigation system, the AV view is shown beside the traditional arthroscope view.

A user study on a knee simulator was done to determine whether the system improves performance in a task of locating particular areas within a knee joint. We hypothesized that the system would improve the performance of novice residents, while doing little to help or hinder the performance of experienced surgeons.

Subjects from novice surgical residents to highly skilled surgeons performed tasks on a knee simulator with and without the aid of the navigation system. Correctness, task completion time, and tool path length were recorded and compared to determine the effectiveness of the system.

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 186–193, 2011.



(a) (b) (c)

Fig. 1. (a) The standard arthroscopic image, which gives few clues to the location of the camera and probe within the joint. (b) The system setup with optical trackers attached to the camera, probe, femur, and tibia. (c) The navigated view showing camera, probe, femur, and tibia. The camera image is projected onto the bones and is visible as the darker area on the femoral condyle.

2 Related Work

In the context of arthroscopic surgery, a number of authors have presented surgical navigation systems that show relative tool positions. But, to the best of our knowledge, none augment the virtual model with live camera images.

Tonet *et al.* [1] described an augmented reality navigation system for knee surgery that shows a 3D model of the patient's anatomy along with the relative tool positions. The field of view of the arthroscope was also dynamically highlighted on the 3D model, but no camera image was shown. No formal user study was made to evaluate the system.

Monahan *et al.* [2] presented a navigation system for arthroscopic hip surgery that tracks tool positions with linkage encoders and displays their relative positions in various 3D views. They performed a user study to determine if their navigation system could help increase speed and accuracy. They found a reduction in the length of the path travelled by the tool and in completion time when using their system.

A number of arthroscopic training systems that include virtual views have also been proposed. Heng *et al.* [3] used virtual models of human anatomy to devise a virtual reality training simulation for arthroscopic surgery. They developed a haptics device to simulate force feedback within a "black box" to present the user with a purely virtual environment. Along with force feedback, their system showed simulated and external views of the arthroscope to aid in training. A user study was not performed to validate their system as a training tool. In a similar way, Bayona *et al.* [4] developed a shoulder arthroscopy training simulator with force feedback. Their interface included a virtual external rendering which showed the viewing cone of the arthroscope, but without live images. They performed user studies to validate the device as a training tool and showed that it was more useful for inexperienced surgeons than for experienced surgeons.

Although augmented virtuality has not been widely used in arthroscopic procedures, a number of authors have developed AV systems for navigation in various other endoscopic procedures.

Paul *et al.* [5] compared AR and AV approaches using a computer system that mixed images from a surgical microscope with 3D preoperative images. They found that the augmented virtuality approach complemented the augmented reality approach because it provided a better understanding of the spatial relationship between the surgical site and the 3D preoperative images.

Dey *et al.* [6] explored the concept of mixed reality to map intra-operative endoscopic video to a patient's anatomy. They fused images of a brain phantom from a tracked endoscope to surfaces derived from 3D preoperative images. Their focus was to paint the video images onto the 3D surface and impart stereoscopic depth cues to provide assistance in surgical planning and guidance. Liao et al. [7] developed a system that fused endoscopic image mosaics with a 3D ultrasound image model to provide extended visualization in intrauterine fetal surgery. Nicolau *et al.* [8] presented a guidance system for laparoscopic surgery which showed both AR and AV views.

The contributions of this paper are (a) an arthroscopic navigation system that projects the camera view onto the joint and (b) a user study showing that the system can improve performance of target localization in novice residents.

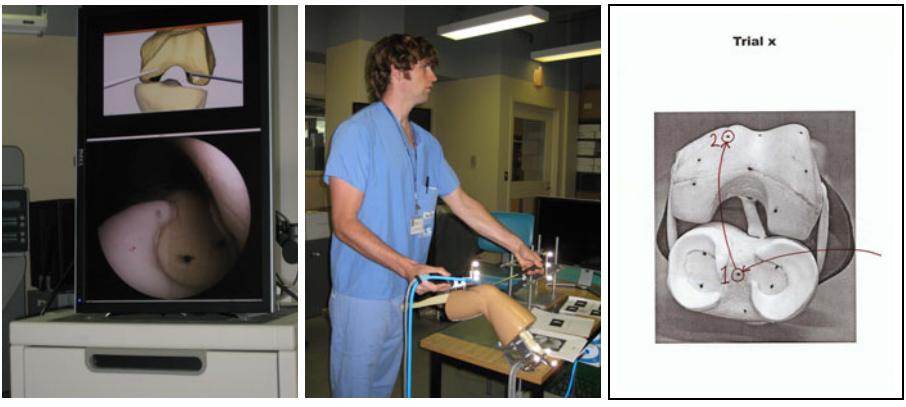
3 System Description

The arthroscope hardware consisted of an IM4000/IM4120 high definition camera system and an HD4300 4mm 30 degree arthroscope from Conmed Linvatec. The knee was modelled with a "Sawbones" artificial knee joint #1413 from Pacific Research Laboratories. The proximal end of the femur was clamped to a table, allowing the tibia to be moved relative to the femur, as is commonly done in surgery.

The components of the system were tracked with a Polaris Hybrid optical tracking system from Northern Digital. Four passive trackers from Traxtal were attached to the arthroscope, the probe, the femur, and the tibia. For the arthroscope and probe, custom mounts were designed and built to hold the trackers. For the femur and tibia, the trackers were held in place by a standard percutaneous clamp.

A thirty inch monitor was positioned in portrait orientation with its bottom edge two meters above the floor, about two meters beyond the knee as seen by the surgeon. (This is a common position for the monitor during arthroscopic surgery.) The standard arthroscopic view was shown on the lower half of the screen. The AV navigation display was in the upper half, showing a 3D computer model of the joint, along with models of the arthroscope and the probe, and the projection of the live arthroscopic image onto the joint (Figure 2a). The AV view was rendered from the same viewpoint, relative to the virtual bones, as the subject saw the plastic bones.

Fiducial markers were fixed to the femur and tibia and a CT scan of those bones was made. Mesh models of the bones were made from the CT scan and the



(a) (b) (c)

Fig. 2. (a) The monitor showing (top) the navigated view and (bottom) the standard arthroscopic image. (b) One of the residents performing a trial. (c) The instruction sheet given for one trial indicating the target points. There are eight target points on the femoral condyles and intracondylar notch (top) and six target points on the tibial plateau (bottom).

locations of the fiducials in the CT coordinate space were recorded. A paired-point algorithm used the fiducial locations to establish the bone-to-tracker transformation. The fiducials were *not* subsequently used as targets in the target localization task.

In order to project the arthroscopic image onto the bone, the OpenCV Computer Vision Library was used to find the intrinsic and extrinsic calibration parameters of the arthroscope camera. OpenGL was used to render a model of the arthroscope, along with the arthroscopic image projected (using GL_REPLACE) onto the virtual bone from the camera position.

The system was limited in that the camera head could not be rotated about its axis, as can be done during surgery, because we did not have the necessary equipment to track the camera head.

4 Experiment Methodology

Eight orthopaedic surgeons and fourteen orthopaedic residents with varying degrees of arthroscopic experience were tested. Each subject performed twelve trials (Figure 2b). In each trial, the subject was asked to locate two target points within the knee joint. The subject would locate the first target point in the arthroscope image, then locate the second target point in the arthroscope image. The sub-millimeter accuracy of the Polaris Hybrid was sufficient for this task. In a variant of the trial, the subject would also touch the probe to each target point.

The trials were performed in four blocks. Each block consisted of three of the same type of trial, but with three different pairs of points. The block order was randomized, as was the trial order within each block. The trials in the four blocks were:

1. Using only the arthroscope image, center each point in the camera view.
2. Using both the arthroscope image and the navigation display, center each point in the camera view.
3. Using only the arthroscope image, center each point in the camera view, then touch it with the probe.
4. Using both the arthroscope image and the navigation display, center each point in the camera view, then touch it with the probe.

The same three pairs of points were used in block 1 and block 2 to allow paired comparison of the results with and without the navigation display. A different set of three pairs was used for both block 3 and block 4. Blocks 1 and 2 were always separated by a different block, as were blocks 3 and 4. In the trials that used only the arthroscope image, the navigation display was turned off.

There were eight target points on the femur and six on the tibia. The locations of the targets were chosen according to a surgeon's description of commonly scoped locations on the femur and tibia. The target points were marked with a black marker on the bone (Figure 2c).

For each trial, the subject was shown a picture of the exposed femur and tibia, with the two target points circled and arrows indicating the order in which they were to be found (Figure 2c). Every trial included one target from the femur and one target from the tibia.

Each subject was shown the layout of the targets on a knee model without the skin covering. The trials were described, and the subject was given ten minutes using the arthroscope and probe to familiarize themselves with the locations of the different targets inside the knee joint. For each trial, we measured the following:

1. Time to locate the target in the camera view.
2. Distance travelled by the camera tip before locating the target.
3. Whether the correct target was found.
4. Time to position the probe on the target (blocks 3 and 4 only).
5. Distance travelled by the probe tip before being positioned on the target (blocks 3 and 4 only).

For the first target, we measured the time and distance from the entry portal on the skin surface until the subject declared that the target was found (even if the wrong target was found). For the second target, we measured the additional time and distance from the first target until the subject declared that the second target was found.

The subject started each trial with the tools in hand and outside of the skin surface. When the subject found a target point, he or she would verbally notify us and we would record the time in the log file, as well as a snapshot of the

Table 1. Experimental results. The bold p values are statistically significant. "NAV" denotes the navigated display, while "STD" denotes the standard display.

Measurement	residents < 2 years N = 8			residents ≥ 2 years N = 6			surgeons N = 8		
	NAV	STD	p	NAV	STD	p	NAV	STD	p
correctness (%)	91.3	79.6	**.002**	90.1	86.1	.130	90.4	86.9	.265
camera time (sec)	24.3	36.6	**.035**	26.7	18.9	**.001**	18.1	17.0	.262
camera distance (cm)	14.0	18.1	.072	14.3	12.1	**.038**	11.6	11.8	.439
probe time (sec)	20.2	24.8	.182	34.1	23.0	**.017**	17.1	19.4	.260
probe distance (cm)	13.8	19.5	**.041**	18.6	14.5	.079	10.8	14.1	.094

arthroscopic image for later verification. The log also continuously recorded the position of both the arthroscope tip and the probe tip, so that we could later calculate the distances travelled. A questionnaire was administered after the trial.

5 Results

The measured experimental results are shown in Table 1. Fifteen hypotheses were tested: For each of the three subject groups and each of the five performance measurements, we performed a t-test to determine whether the performance measurement using the navigated display was superior to that measurement using only the standard display.

For residents with less than two years of experience, there was a **significant improvement** in acquiring the correct target (91.3% versus 79.6%, p = 0.002), in the time taken to locate the target with the camera (24.3 seconds versus 36.6 seconds, p = 0.035), and in the distance travelled by the probe to touch the target (13.8 cm versus 19.5 cm, p = 0.041).

For residents with at least two years of experience, there was a **significant worsening** in the time to locate the target with the camera (26.7 seconds versus 18.9 seconds, p = 0.001), in the distance travelled by the camera to locate the target (14.3 cm versus 12.1 cm, p = 0.038), and in the time for the probe to touch the target (34.1 seconds versus 23.0 seconds, p = 0.017).

For surgeons, no significant differences were found between the navigated and standard methods.

From the questionnaires, which gathered responses on a seven-point Likert scale:

- Subjects at all three levels agreed that the navigated display gave a better understanding of the 3D configuration of the bone, tool, and camera.
- Subjects were neutral or agreed that it was easier to find and touch the fiducials using the navigated display instead of only the traditional arthroscopic image.

- Novice residents and surgeons were neutral that the lack of camera rotation made the task more difficult, while experienced residents strongly agreed.
- Surgeons agreed that the experimental setup accurately simulated what they would experience in the operating room, while residents, as a group, were neutral.

6 Discussion

The navigated display is clearly beneficial to residents with less than two years of arthroscopic experience. This strongly supports our hypothesis that the navigated display can assist early residents.

The navigated display is clearly detrimental to residents of at least two years of arthroscopic experience. We believe that this is because those residents have already established a mental model of the arthroscopic environment which is different than the 3D model that the AV navigation system provided. In fact, one experienced resident explicitly said that his mental model clashed with that of the navigated display.

We could not find any significant effect of the navigated display upon surgeons, all of whom had at least five years of arthroscopic experience as surgeons. We believe that this is because the surgeons were so highly trained that they did not need, and hence paid little attention to, the navigated display. However, five of the eight surgeons noted that they occasionally used the display to confirm the location of target points that were harder to find, so there is potential for intraoperative use by surgeons. In addition, six out of eight surgeons commented that the system would be beneficial as a training tool for novice residents.

The experimental setup seems to be close to what would be experienced in surgery, according to the experienced surgeons. But the lack of camera rotation made the task more difficult than would be encountered in real surgery.

All subjects believed that the navigated display gave a better understanding of the 3D configuration of the bone, tool, and camera. Although no group of subjects strongly believed that the navigated display made the target points easier to locate, we measured a substantial performance improvement among the novice residents. In our observations and interviews of the novice and experienced residents, we found that they used the navigated display to locate the general area of the target, then used the traditional arthroscopic view to adjust their final position. That suggests that the live camera view projected onto the bone may not be necessary.

7 Conclusion

The study shows strong evidence that the AV navigated display can assist novice residents in the difficult task of locating anatomical locations in arthroscopic surgery.

The contrary results from experienced residents suggest that many of those residents form a mental model of the procedure that is different from our 3D

model. Further study should be done to determine this trained mental model. Either our 3D model could be adapted to the trained model, or another study could be performed to compare the utility of the trained model to that of our 3D model.

The most important improvement in the navigated system would be to incorporate camera rotation, although the surgeons of our study agreed that our system accurately reflected what would be experienced in the operating room.

Acknowledgements. Thank you to our subjects for their participation and to the reviewers for their helpful comments. This work is supported by Natural Sciences and Engineering Research Council of Canada (NSERC) grant STPGP 336797 and by the joint NSERC/Canadian Institutes of Health Research CHRP grant 385959.

References

1. Tonet, O., Megali, G., D'Attanasio, S., Dario, P., Carrozza, M.C., Marcacci, M., Martelli, S., La Palombara, P.F.: An augmented reality navigation system for computer assisted arthroscopic surgery of the knee. In: Delp, S.L., DiGoia, A.M., Jaramaz, B. (eds.) MICCAI 2000. LNCS, vol. 1935, pp. 1158–1162. Springer, Heidelberg (2000)
2. Monahan, E., Shimada, K.: A study of user performance employing a computer-aided navigation system for arthroscopic hip surgery. International Journal of Computer Assisted Radiology and Surgery 2(3-4), 245–252 (2007)
3. Heng, P.A., Cheng, C.Y., Wong, T.T., Xu, Y., Chui, Y.P., Chan, K.M., Tso, S.K.: A virtual-reality training system for knee arthroscopic surgery. IEEE Transactions on Information Technology in Biomedicine 8(2), 217–227 (2004)
4. Bayona, S., Fernández-Arroyo, J.M., Martín, I., Bayona, P.: Assessment study of insight ARTHRO VR arthroscopy virtual training simulator: face, content, and construct validities. Journal of Robotic Surgery 2(3), 151–158 (2008)
5. Paul, P., Fleig, O., Jannin, P.: Augmented virtuality based on stereoscopic reconstruction in multimodal image-guided neurosurgery: methods and performance evaluation. IEEE Transactions on Medical Imaging 24(11), 1500–1511 (2005)
6. Dey, D., Gobbi, D.G., Slomka, P.J., Surry, K.J.M., Peters, T.M.: Automatic fusion of freehand endoscopic brain images to three-dimensional surfaces: creating stereoscopic panoramas. IEEE Transactions on Medical Imaging 21(1), 23–30 (2002)
7. Liao, H., Tsuzuki, M., Mochizuki, T., Kobayashi, E., Chiba, T., Sakuma, I.: Fast image mapping of endoscopic image mosaics with three-dimensional ultrasound image for intrauterine fetal surgery. Minimally Invasive Therapy & Allied Technologies 18(6), 332–340 (2009)
8. Nicolau, S.A., Goffin, L., Soler, L.: A Low Cost and Accurate Guidance System for Laparoscopic Surgery: Validation on an Abdominal Phantom. In: Proceedings of the ACM Symposium on Virtual Reality Software and Technology, pp. 124–133. ACM, New York (2005)

Bronchoscopy Navigation beyond Electromagnetic Tracking Systems: A Novel Bronchoscope Tracking Prototype

Xiongbiao Luo[1], Takayuki Kitasaka[2], and Kensaku Mori[3,1]

[1] Graduate School of Information Science, Nagoya University, Japan
[2] Faculty of Information Science, Aichi Institute of Technology, Japan
[3] Information and Communications Headquarters, Nagoya University, Japan

Abstract. A novel bronchoscope tracking prototype was designed and validated for bronchoscopic navigation. We construct a novel mouth- or nose-piece bronchoscope model to directly measure the movement information of a bronchoscope outside of a patient's body. Fusing the measured movement information based on sequential Monte Carlo (SMC) sampler, we exploit accurate and robust intra-operative alignment between the pre- and intra-operative image data for augmenting surgical bronchoscopy. We validate our new prototype on phantom datasets. The experimental results demonstrate that our proposed prototype is a promising approach to navigate a bronchoscope beyond EMT systems.

1 Introduction

To develop bronchoscopic navigation, accurate and stable alignment between a pre-built 3D anatomical airway model and real-time 2D bronchoscopic video frames remains challenging. Although EMT-based methods provide global synchronization between patients and their CT coordinate systems [1,2], such techniques are still constrained due to their disadvantages such as sensitivity to localization problems caused by patient movement and inaccurate measurements resulted from magnetic field distortion. Image-based schemes usually register the world and CT coordinate systems by calculating the similarities between real bronchoscopic video frames and virtual bronchoscopic images generated from CT-derived virtual bronchoscopy [3,4]. Since image-based techniques depend on local texture information such as bifurcations or folds, they usually provide accurate local registration between patients and their CT coordinate systems. However, such approaches fail easily to track the global trajectory of a bronchoscope due to uncertainties (e.g., bubbles) that commonly occur in bronchoscopy.

This work explores a novel bronchoscope prototype that deals with the difficulties of EMT- and image-based methods during bronchoscopic navigation. Our main contributions are summarized as follows. First, a novel bronchoscope tracking prototype was constructed and evaluated on a bronchial phantom. We originally proposed a novel external tracking method on the basis of an optical mouse (OM) sensor and demonstrated its effective and promising performance

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 194–202, 2011.

for bronchoscope motion estimation. Compared to our previous work [5], this study shows the complete idea of our tracking prototype and more attractive tracking performance by combining SMC methods. Next, currently, although only 2D motion information of a bronchoscope can be measured by an OM sensor in our proposed prototype, such an OM sensor-based bronchoscope mockup still has several advantages in contrast to other external tracking methods, such as EMT-based. These advantages include: (1) constructing such a bronchoscope tracking model is very cheap and simple; (2) since we design the OM sensor and the rotary encoder to be located outside a patient's airway tree, such a mockup does not occupy the space of a bronchoscope tip or its working channel but the EMT system currently requires an EMT sensor to be attached at either the surface or inside the working channel of the bronchoscope tip; this constrains the movement of the bronchoscope in big bronchial branches; and (3) the OM sensor and encoder measurements are unaffected by ferromagnetic metals or conductive material within the bronchoscope; they usually distort the magnetic fields of EMT systems and cause inaccurate measurements of EMT sensors. Third, sometimes image-based approaches unavoidably misalign patients and their CT coordinates because of inherent global uncertainties, e.g., image artifacts. After a failure to register real and virtual images for several frames, an image-based algorithm usually cannot automatically recover the tracking procedure by itself due to shortages of the global insertion depth and rotation information around the running direction of the bronchoscope. We solve such a problem using an OM sensor to directly measure such global motion information of the bronchoscope and reduce opportunities to get trapped in local minima during bronchoscope tracking. Fourth, to incorporate OM sensor output for bronchoscope tracking, we introduce SMC sampler [6], which proved to be an effective means to combine different tracking sources since it can somewhat tackle situations where ambiguities occur in bronchoscopic videos [7]. Finally, we also believe our novel bronchoscope model can be conveniently and easily integrated into intra-bronchoscopy inside operating rooms without any overloads since we can attach a rotary encoder on the angle lever surface and design to fix an OM sensor on a mouth- or nose-piece that is also indispensable for a conventional bronchoscope.

2 Novel Bronchoscope Tracking Prototype

2.1 Prototype Overview

The movements of a bronchoscope are usually comprised of three parts: (1) moving it inside or outside airway trees along the running direction, (2) rotating it around the running direction, and (3) bending its bendable section by its angle lever fixed on its control head. Three parameters corresponding to three motion components are introduced to characterize the bronchoscope movements: the insertion depth inside the airway trees, rotational angle around the running direction, and the bending angle of the bendable section.

Based on these movement properties of a bronchoscope, we designed our prototype with three key functions that are clarified as follows: (1) it directly measures

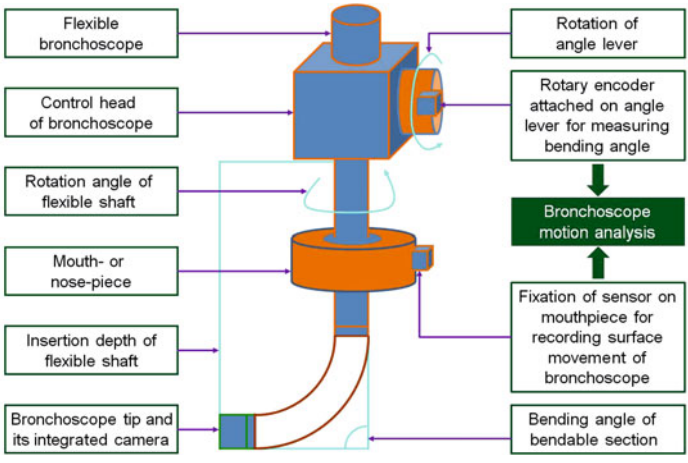


Fig. 1. Our designed bronchoscope prototype with a sensor fixed on a mouth- or nose-piece to measure insertion depth and rotation information along the running direction of a bronchoscope and a rotary encoder attached at angel lever surface to determine the bending angle of the bendable section during interventions

the insertion depth of the bronchoscope shaft, (2) it immediately records the rotational angle around the viewing (running) direction of the bronchoscope camera, (3) it automatically obtains the rotational information of the angle lever to determine the bending angle of the bendable section. Note that this movement information is acquired outside of a patient's body, and no additional sensors are attached at the bronchoscope shaft surface or its working channel. Fig. 1 outlines the prototype with different components relative to different functions.

2.2 Bronchoscope Motion Analysis

A. Pre-processing
Since unavoidable time-delay occurs between 2D OM sensor measurements (sample rate: 2000∼6400 fps) and bronchoscopic images (frame rate: 30 fps) during data collection, we must synchronize the OM sensor outputs and video frames based on their timestamps. If the timestamps represent the exact time of two kinds of outputs, we can calibrate them through a linear interpolation.

B. Fusion of OM Sensor Measurements
After temporal calibration between the OM sensor outputs and the bronchoscopic frames, we obtain relative insertion depth Δ and rotational angle Ψ of the running direction of the bronchoscope between successive bronchoscopic images. To predict six DOF parameters of bronchoscope motions, we incorporate measurements Δ and Ψ into the SMC simulation because it proved to be an effective means to fuse different external tracking sources.

Sequential Monte Carlo samples were deterministically drifted and stochastically diffused to predict the posterior probability density of interest. We here utilize an SMC simulation approach that originated from the work in [6].

We produce set of random samples $\mathcal{S}_i^k = \{(\mathbf{x}_i^k, w_i^k, c_i^k) : i = 1, 2, 3, \ldots, N; k = 1, 2, 3, ..., M\}$ (N and M are the number of frames and samples, respectively, M is set to 500 since it met with a compromise between accuracy and computational complexity in our case; w_i^k is a sample weight; c_i^k denotes the accumulative weight of each sample) to approximate the posterior probabilistic distribution of current bronchoscope camera motion parameters $\mathbf{x}_i^k$, which are defined on the basis of transformation matrix ${}^{CT}\mathbf{T}_C^{(i)}$ including translation ${}^{CT}\mathbf{t}_C^{(i)}$ and rotation ${}^{CT}\mathbf{R}_C^{(i)}$ from the bronchoscope camera coordinate system to the CT coordinate system at time i. Then our proposed bronchoscope tracking method using SMC simulation is performed in the following two steps:

State Transmission. The new state of each sample can be determined by deterministic drift and stochastic diffusion by transform function F:

$$\mathbf{x}_i^k = F(A\mathbf{x}_{i-1}^k, Bn_i^k), \tag{1}$$

where A denotes a deterministic inter-frame motion and depends on the OM sensor measurements (or observations) of relative insertion depth Δ and rotational angle Ψ of the running direction, B describes an uncertainty or a stochastic part of the relative motion in Eq. 1, and n_i^k is an independent stochastic variable.

During this state dynamic step, SMC simulation requires a probabilistic density function to present state transmission probability $p(\mathbf{x}_i^k|\mathbf{x}_{i-1}^k)$ between consecutive time steps. Since we have no prior knowledge of the bronchoscope camera movement, in other words, since we do not know prior probabilistic distribution $p(\mathbf{x}_i)$ for state vector $\mathbf{x}_i$, we employ a random walk on the basis of normal distribution with respective to noise n_i^k: $n_i^k \sim \mathcal{N}(\mu, \sigma^2)$ to approach dynamic density $p(\mathbf{x}_i^k|\mathbf{x}_{i-1}^k)$ [6,8]:

$$p(\mathbf{x}_i^k|\mathbf{x}_{i-1}^k) \propto \frac{1}{\sqrt{2\pi}\sigma} \exp(-(B^{-1}(\mathbf{x}_i^k - A\mathbf{x}_{i-1}^k) - \mu)^2/2\sigma^2). \tag{2}$$

Note that A and Bn_i^k need to be determined in Eq. 1. For A, according to relative insertion depth Δ and rotational angle Ψ, we compute the relative motion with the translation vector $\Delta\hat{\mathbf{t}}_k^{(i)}$ and the Euler rotation angles $\Delta\hat{\alpha}_k^{(i)}$ as:

$$\Delta\hat{\mathbf{t}}_k^{(i)} = [0,\ 0,\ \Delta]^T, \quad \Delta\hat{\alpha}_k^{(i)} = [0,\ 0,\ \Psi]. \tag{3}$$

By transforming the rotation part from Euler angles to matrix, the deterministic drift part A can be determined by

$$A = \begin{pmatrix} \Delta^{CT}\mathbf{R}_C^{(i)}\left(\Delta\hat{\alpha}_k^{(i)}\right) & \Delta\hat{\mathbf{t}}_k^{(i)} \\ \mathbf{0}^T & 1 \end{pmatrix}. \tag{4}$$

Additionally, the stochastic diffusion component Bn_i^k can be computed by Eq. 4 in terms of $\Delta\tilde{\mathbf{t}}_k^{(i)} = \left[n_t^k,\ n_i^k t,\ n_i^k t\right]^T$ and $\Delta\tilde{\alpha}_k^{(i)} = \left[n_i^k\theta,\ n_i^k\theta,\ n_i^k\theta\right]$, where t and θ are translational and rotational constants during stochastic diffusion.

Observation Density. After state evolution, we must compute sample weight w_i^k and observation density $p(\mathbf{y}_i|\mathbf{x}_i)$ ($\mathbf{y}_i$ is an observation variable). In this work, w_i^k is defined as the modified mean squared error ($MoMSE$) similarities between real bronchoscopic frame $\mathbf{I}_R^{(i)}$ and virtual image $\mathbf{I}_V$ produced from estimate $\mathbf{x}_i^k$ using volume rendering techniques; it can be presented by [4]:

$$w_i^k = MoMSE(\mathbf{I}_R^{(i)}, \mathbf{I}_V(\mathbf{x}_i^k)). \tag{5}$$

Based on factored sampling scheme [6], observation density $p(\mathbf{y}_i|\mathbf{x}_i)$ can be calculated by

$$p(\mathbf{y}_i|\mathbf{x}_i = \mathbf{x}_i^k) = w_i^k(\sum_{j=1}^{M} w_i^j)^{-1}. \tag{6}$$

Finally, the pose parameters of the current bronchoscope and its combined camera pose can be determined in terms of sample weight w_i^k:

$$\tilde{\mathbf{x}}_i = \max_{\tilde{w}_i^k}\{(\mathbf{x}_i^k, w_i^k)\}. \tag{7}$$

The final estimate corresponds to motion state $\tilde{\mathbf{x}}_i$ with weight $\tilde{w}_i$ to maximize the similarity between the current real and virtual frames.

3 Experimental Results

Since we currently have no patient data for our new bronchoscope tracking device, we validate the proposed tracking prototype on a phantom. We investigate three tracking schemes: (1) Deguchi et al. [4], only image registration using similarity measure $MoMSE$, (2) Schwarz et al. [1], using an EMT system, and (3) our new method presented in Section 2. To evaluate the tracking accuracy of the three methods, we generate two sets of ground truth data (GTD) using manual registration to align the RB and virtual bronchoscopic (VB) viewing points by hand. We then calculated the position and orientation errors by $\delta = \|\mathbf{t} - \mathbf{t}_G\|$, $\phi = \arccos((trace(\mathbf{R}\mathbf{R}_G{}^T) - 1)/2)$, where δ and ϕ are Euclidean distance and rotation error around the invariant Euler axis in accordance with estimated pose $(\mathbf{R} \mid \mathbf{t})$ and reference (ground truth) pose $(\mathbf{R}_G \mid \mathbf{t}_G)$ [9], where $\mathbf{t}$ and $\mathbf{t}_G$ denote translation, $\mathbf{R}$ and $\mathbf{R}_G$ are rotation matrices.

Table 1 summarizes the position and orientation errors by contrasting GTDs with the tracking results from the three approaches. The average position errors of the three methods were 25.5 mm, 5.09 mm and 1.10 mm. Simultaneously, the average orientation error was 3.88° by the proposed method, compared to 37.0° and 11.1° from the other two approaches. Fig. 2 illustrate the tracking accuracy of the predicted results of Experiments B in contrast to the ground

Table 1. Examples of position and orientation errors in tracking results using methods of Deguchi et al. [4], Schwarz et al. [1], and our new method

GTD tests	Position error (mm) and orientation error (degrees)											
	Deguchi et al. [4]				Schwarz et al. [1]				Our new method			
	Position		Orientation		Position		Orientation		Position		Orientation	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
A	32.8	37.2	42.9	50.5	5.70	2.77	12.2	3.53	1.40	3.19	5.20	11.3
B	18.2	23.9	31.0	47.1	4.48	2.66	10.0	3.38	0.80	0.50	1.68	1.37
Ave.	**25.5**	**30.6**	**37.0**	**48.8**	**5.09**	**2.72**	**11.1**	**3.50**	**1.10**	**1.85**	**3.88**	**6.20**

Table 2. Comparison of registered results based on visual inspection

Phantom Experiments	Number of frames	Number (%) of successfully processed frames		
		Deguchi et al. [4]	Schwarz et al. [1]	Our new method
A	1556	763 (49.0%)	708 (45.5%)	1387 (89.1%)
B	1498	832 (55.5%)	1001 (66.8%)	1462 (97.6%)
C	1183	83 (7.02%)	948 (80.1%)	966 (81.7%)
D	1805	198 (11.0%)	1310 (72.6%)	1582 (87.7%)
E	1032	416 (40.3%)	538 (52.1%)	639 (61.9%)
F	1228	606 (49.5%)	689 (56.1%)	1106 (90.1%)
Total	**8302**	**2898 (34.9%)**	**5194 (62.6%)**	**7142 (86.0%)**

truth. Our new tracking prototype is more accurate and stable than the other two. Moreover, according to visual inspection, successfully processed RB frames were quantified in Table 2. Our method successfully processed a total of 7142 (86.0%). Fig. 3 displays the RB and VB images at the selected frames by all three methods. Both prove the better accuracy and robustness of our model.

4 Discussion

The contributions of this work were already clarified in Section 1. Our method provides the insertion depth as global information to calculate the image similarity and hence improve its robustness at the branching cases that will be further addressed in the future. Compared to EMT-based approaches, our method will never be involved in any inherent calibration or registration such as camera-sensor calibration and CT-to-physical space registration that inevitably introduce many errors before tracking with the bronchoscope. Even a bronchoscope itself contains conductive metals that distort the magnetic field and deteriorate EMT accuracy; our prototype remains free of such a problem. In contrast to image-based algorithms, our proposed method directly provides global information on the insertion depth and rotational angle of the viewing direction to predict bronchoscope motions combined with SMC simulation without any optimization procedures that get easily trapped in local minima due to the shortages of global information and such image artifacts as bubbles or patient movements.

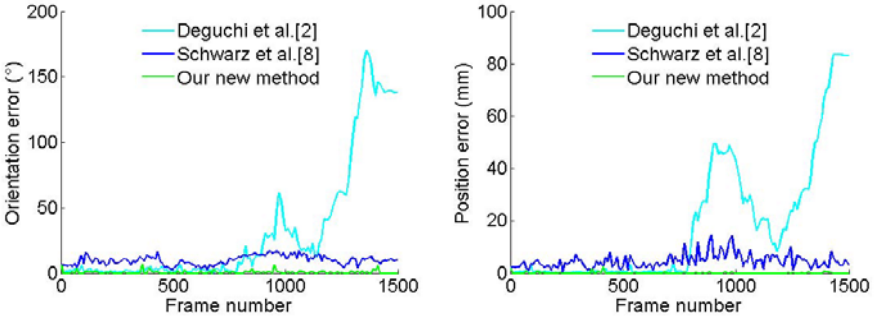


Fig. 2. Orientation error (*left*) and position error (*right*) of Experiment B are plotted from tracking results predicted by methods of Deguchi et al. [4] (*cyan*), Schwarz, et al. [1] (*blue*), and our new method (*green*).

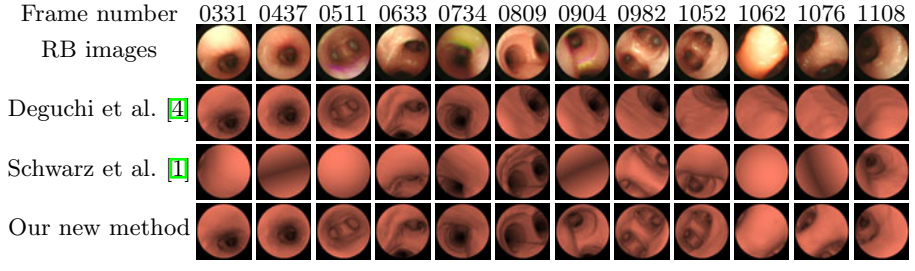


Fig. 3. Visual comparison of tracking results of Experiment F using different methods during phantom validation. Top row shows selected frame numbers, and second row shows their corresponding phantom RB images. Other rows display virtual bronchoscopic images generated from tracking results using the methods of Deguchi et al. [4], Schwarz et al. [1], and our new method. Our proposed method shows the best performance.

However, in our experiments, the proposed method still misaligns some RB and VB frames when continuously tracking bronchoscope motions during navigated bronchoscopy. One main reason behind these misalignments might be the shortage of the bending angle information of the bronchoscope bendable part. We did not realize the function of automatically measuring the bending angle. We currently only provide information on the insertion depth and the rotation of the running direction for motion prediction. Otherwise, the nonlinear illumination changes result in dark video images; hence, they cause difficulties of sample weight calculation, and incorrect sample weights unavoidably happen.

Moreover, the following issues must be clarified. First, although our method provides significant tracking accuracy, we clarified that the position and orientation errors shown in Table 1 are stated relative to GTD. However, GTD itself involves errors originating from our manual registration or synchronization.

Next, since bronchoscopic images were collected inside a static phantom without respiration motion or other deformations, another particular challenge of airway deformation was not explicitly validated in this work. We need to further evaluate our tracking prototype on patient data. Additionally, the computational efficiency of the proposed method is about 1.1 seconds per frame without speed optimizations and multi-threading. Last, although we concentrated on developing bronchoscope motion tracking for navigated bronchoscopy, our prototype should also be appropriate to navigate other endoscopes, e.g., colonoscope.

5 Conclusions and Future Work

This paper proposed a novel bronchoscope tracking prototype and its validation on phantom datasets. We realized two functions of our model and applied them to estimate bronchoscope movements. The experimental results demonstrated the accuracy and robustness of our method by pose errors of only 1.10 mm and 3.88°, successfully registered a total of 7142 (86.0%) images, and increased the tracking performance by at least 23.4%, compared to image- and EMT-based methods. Besides evaluation of our method on patient data and reducing computational times, future work also includes the extension of another function of our bronchoscope prototype and further improvement of its performance.

Acknowledgments. This work was supported in part by Grant-in-Aid for Science Research funded by MEXT and JSPS, and "Chi-no-kyoten" by Aichi Prefecture. Authors would like to thank Dr. H. Tahabatake for providing datasets.

References

1. Schwarz, Y., Greif, J., Becker, H.D., Ernst, A., Mehta, A.: Real-time electromagnetic navigation bronchoscopy to peripheral lung lesions using overlaid CT images: the first human study. Chest 129(4), 988–994 (2006)
2. Soper, T.D., Haynor, D.R., Glenny, R.W., Seibel, E.J.: In vivo validation of a hybrid tracking system for navigation of an ultrathin bronchoscope within peripheral airways. IEEE TBME 57(3), 736–745 (2010)
3. Deligianni, F., Chung, A.J., Yang, G.Z.: Nonrigid 2-D/3-D registration for patient specific bronchoscopy simulation with statistical shape modeling: Phantom validation. IEEE TMI 25(11), 1462–1471 (2006)
4. Deguchi, D., Mori, K., Feuerstein, M., Kitasaka, T., Maurer Jr., C.R., Suenaga, Y., Takabatake, H., Mori, M., Natori, H.: Selective image similarity measure for bronchoscope tracking based on image registration. MedIA 13(4), 621–633 (2009)
5. Luo, X., Feuerstein, M., Kitasaka, T., Mori, K.: A novel bronchoscope tracking method for bronchoscopic navigation using a low cost optical mouse sensor. In: Wong, K.H., Holmes, D.R. (eds.) SPIE Medical Imaging 2011, Florida USA, vol. 7964, pp. 79641T (2011)
6. Isard, M., Blake, A.: Condensation - conditional density propagation for visual tracking. International Journal of Computer Vision 29(1), 5–28 (1998)

7. Luo, X., Reichl, T., Feuerstein, M., Kitasaka, T., Mori, K.: Modified hybrid bronchoscope tracking based on sequential monte carlo sampler: Dynamic phantom validation. In: Kimmel, R., Klette, R., Sugimoto, A. (eds.) ACCV 2010, Part III. LNCS, vol. 6494, pp. 409–421. Springer, Heidelberg (2011)
8. Arulampalam, M., Maskell, S., Gordon, N., Clapp, T.: A tutorial on particle filters for nonlinear/non-gaussian Bayesian tracking. IEEE TSP 50(2), 174–188 (2002)
9. Schneider, M., Stevens, C.: Development and testing of a new magnetic-tracking device for image guidance. In: Cleary, K.R., Miga, M.I. (eds.) SPIE Medical Imaging 2007, California USA, vol. 6509, pp. 65090I (2007)

3D Tracking of Laparoscopic Instruments Using Statistical and Geometric Modeling*

Rémi Wolf[1], Josselin Duchateau[1], Philippe Cinquin[1], and Sandrine Voros[2]

[1] UJF-Grenoble 1, CNRS
[2] UJF-Grenoble 1, CNRS, INSERM
TIMC-IMAG UMR 5525, Grenoble, F-38041, France

Abstract. During a laparoscopic surgery, the endoscope can be manipulated by an assistant or a robot. Several teams have worked on the tracking of surgical instruments, based on methods ranging from the development of specific devices to image processing methods. We propose to exploit the instruments' insertion points, which are fixed on the patients abdominal cavity, as a geometric constraint for the localization of the instruments. A simple geometric model of a laparoscopic instrument is described, as well as a parametrization that exploits a spherical geometric grid, which offers attracting homogeneity and isotropy properties. The general architecture of our proposed approach is based on the probabilistic Condensation algorithm.

Keywords: laparoscopic surgery, image-based localization of surgical instruments, Condensation algorithm.

1 Introduction

Minimally invasive surgery (MIS) has become more and more popular with surgeons and the public in the recent years, although its superiority on open surgery has not been systematically proven in terms of per and post-operative complications ([8]). Furthermore, the mini-invasive approach is more expensive and more challenging technically for the surgeons than open surgery (loss of depth information, loss of tactile information, limited field of view, inverted movements of the instruments due to their insertion in the patient's abdominal cavity through "fixed" points on the abdominal wall).

These limitations highlight the need to develop further new methodologies dedicated to MIS. These methodologies can be divided in two categories: skills (learning, practicing and evaluating) and tools to perform these skills (end-effectors, endoscopic cameras). New approaches to learn MIS have already been developed, especially using virtual reality simulators [7] and video recordings of surgeries for the evaluation of the surgeon's performance [1]. Extensive research has also been performed on the optimization of laparoscopic tools, the most outstanding being the development of robots, either holding the endoscope inside

* This work has been supported by French National Research Agency (ANR) through TecSan program (project DEPORRA nANR-09-TECS-006).

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 203–210, 2011.

the abdominal cavity (AESOP®, ViKY®) or offering more degrees of freedom than the standard instruments (DaVinci®). Here we will concentrate on the enhancement of the endoscope manipulator. It is bothersome and distracting for the surgeons to control the positioning of the robot manually or by voice. Therefore there has been some research to visually track the tip of the surgeon's instruments and adapt the position of the endoscope to follow them[3,9].

In this paper, we present a new method to localize a laparoscopic instrument from video images in real time, using only a priori knowledge on the geometry of the instrument, the position of its insertion point and a statistical model for the displacement of the instrument based on the Condensation algorithm. Compared to other methods described until today, this algorithm allows us to automatically retrieve the 3D orientation and tip position of the instrument in the camera referential, offering new opportunities such as analysing the motion of the instruments to classify the operating task.

2 Materials and Methods

2.1 General Framework

The goal of our method is to track the motion of a laparoscopic instrument inside the abdominal cavity without modifying it (ie. without a physical tracking system such as infrared or magnetic localizers), using only information from the endoscopic image and a geometric model of the tool.

After an initial calibration procedure, we determine the insertion point of each instrument in 3D using a priori knowledge about its geometry. We then build a geode centered on the insertion point describing the different orientations that can be taken by the tool. A propagation algorithm using particles evolving over the geode's elements according to a determinist and a stochastic law is then computed (the **Con**ditional **Dens**ity Propag**ation** or Condensation algorithm [4]), with measures performed at each time step to guess the most likely orientation of the tool, after reconstruction and projection of the instrument model in the 2D image. Finally, we determine the tip depth along this orientation with an Otsu-based [5] segmentation method over a sliding window.

2.2 System Calibration

Our method requires system calibration: in particular it must be possible to calculate the 3D line in the fixed field of reference (FOR) that corresponds to a 2D image point on the screen. This requires an intrinsic calibration and distortion correction performed using Zhang's procedure[11] with a planar chessboard, which is validated in the medical imaging community.

If the endoscope is displaced by a robot, it is also necessary to know the camera's displacement in the robot's referential, which requires a "hand-eye" calibration of the robotic system. In this paper, we choose to work with a fixed camera, since we concentrate on the presentation of our localization method.

2.3 Instrument Model

Laparoscopic tools can have differents appearances, according to their color (which is not standardized) or to their tip which varies according to their fonction. However, they all have in common a cylindrical shape and a standardized radius, in order to fit in trocars. It is thus possible to model a tool as a finite cylinder of known diameter [2]. Our search, given a predefined insertion point (see section 2.4), is thus limited to its two angular parameters (pan and tilt) that define a direction axis, depicted by a 3D unit vector $\mathbf{e_1}$, and a translational parameter, γ, representing the instrument's depth along its axis (see section 2.6).

The knowledge of the axis of the tool and of its insertion point permits the reconstruction of its borders as seen by the camera (see fig.1, left): the vector $\mathbf{CT}$ is the translation vector representing the instrument insertion point T in the camera's FOR (where C is the camera's center of projection). We can construct a referential linked to the tool by defining two new vectors as follow:

$$\mathbf{e_2} = \frac{\mathbf{T} \otimes \mathbf{e_1}}{\|\mathbf{T} \otimes \mathbf{e_1}\|} \ , \quad \mathbf{e_3} = \mathbf{e_1} \otimes \mathbf{e_2}$$

The instrument axis can be defined by the line (T,$\mathbf{e_1}$) but also by the line (P,$\mathbf{e_1}$), where $\mathbf{CP} = \|\mathbf{T} \otimes \mathbf{e_1}\|\mathbf{e_3}$, as shown in fig. 1, left. The plane (P,$\mathbf{e2}$,$\mathbf{e3}$) is the only right section of the cylinder representing the tool that contains the camera's center of projection C. Working in this right section (fig.1, right), its is possible to define two points S_1 and S_2 that belong to the tangent planes to the instrument running through the camera center. S_1 and S_2 satisfy $\|\mathbf{PS}_{1,2}\| = \rho$, where ρ is the instrument's radius and $\mathbf{CS}_{1,2} \cdot \mathbf{PS}_{1,2} = 0$, i.e. the following system:

$$\begin{cases} \|\alpha\mathbf{e_2} + \beta\mathbf{e_3}\| = \rho \\ (\alpha\mathbf{e_2} + (\|\mathbf{T} \otimes \mathbf{e_1}\| + \beta)\mathbf{e_3}) \cdot (\alpha\mathbf{e_2} + \beta\mathbf{e_3}) = 0 \end{cases} \quad (\alpha, \beta) \in \mathbb{R}^2$$

This system admits two solutions (one for each tangent plane) :

$$(\alpha, \beta) = \left(\pm\rho\sqrt{1 - \frac{1}{\|\mathbf{T} \otimes \mathbf{e_1}\|}}, -\frac{\rho^2}{\|\mathbf{T} \otimes \mathbf{e_1}\|} \right)$$

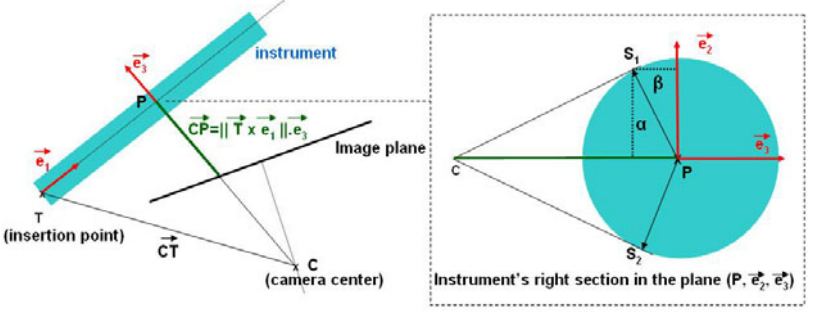


Fig. 1. Description of the tool model

Once these parameters have been extracted, one can easily build in 3D the lines defined by the tangent plane to the instrument passing through the camera center (corresponding to the instrument borders in the 2D image).

$$L = \mathbf{T} \pm |\alpha|\mathbf{e_2} + \beta\mathbf{e_3} + \gamma\mathbf{e_1} \quad \gamma \in \mathbb{R}$$

2.4 Determination of the Insertion Point

Several methods have already been developed in order to retrieve the lines representing the borders of a laparoscopic instrument in 2D images[9,3]. Here we use manual selection of these lines in a sequence of images with one camera position and a moving instrument, which allows us to find (thanks to the calibration and the geometric model of the tool) a beam of 3D lines each representing an instrument's axis. The intersection of these lines, which corresponds to the insertion point, is computed by solving a rank-3 system with a SVD decomposition. We will see in the discussion that this method could easily be automatized.

2.5 Instrument Localization

Our tracking system is based on the Condensation algorithm[4]: it uses factored sampling in which randomly generated sets represent the probability distribution of a certain feature. The random set is propagated using a dynamic model and visual observations, which is of particuliar interest in our situation in which background clutter can often cause direct observation to fail. We therefore need to build a specific dynamic model for our process, along with the corresponding observation model. Given our model construction, it is preferable to detect the instrument 3D position in two steps, determining the angular parameters before the translational parameter, since this parameter only influences the length of the apparent segments, and not their orientation.

Geodesic Geometry. Two parameters are needed to describe the angular orientation of the tool in a 3D space. The typical representation of these parameters would be standard latitude and longitude coordinates (pan and tilt), but these are not well suited to our problem as they are not isotropic (non-uniform distribution) and computational stability near the poles is non-trivial. A solution developed in the field of climate modeling is the use of quasi-uniform spherical geodesic grids [6] among which icosahedral grids give almost homogeneous and quasi-isotropic coverage of the sphere. These grids are obtained by the successive subdivisions of an isocahedron (fig. 2a). Parametrization is easily obtained by dividing the geodesic sphere into 5 panels (fig. 2b) composed almost exclusively of hexagonal cells. Each panel, once rotated and twisted, can be represented as a rectangular array (fig. 2c).

Each 3D unit vector that runs from the center of the sphere to a cell on its surface corresponds to a unique cell on one of the 5 rectangular arrays, as illustrated by the hexagonal cell marked in blue on the sphere in fig. 2a, which corresponds to an unique cell of the rectangular array in fig. 2c. This vector can easily be computed given the 2D position of the cell in the rectangular array, and in our algorithm

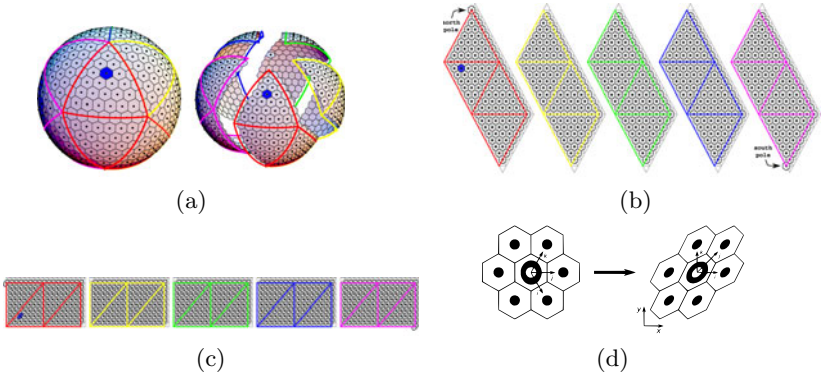


Fig. 2. Representation of the geodesic grid: (a) The spherical geodesic grid composed of hexagonal and pentagonal cells. (b) Cutting the geodesic sphere in 5 panels. (c) Twisted and rotated panels can be represented in rectangular arrays. (d) A particle evolves freely in three independent directions in the hexagonal tiling. Note that $j = i + k$. (images source http://kiwi.atmos.colostate.edu/BUGS/geodesic)

correspondences between 2D cells and 3D vectors are computed and stored in correspondence tables in order to increase speed for the rest of the execution.

Particle Model. In our model, each particle is described at a given time by five variables: the current and previous 2D coordinates of the particle $\mathbf{s_t} = [x_t, y_t, x_{t-1}, y_{t-1}]$ expressed in the panel variable p_t.

Particle dynamics are controlled by a linear stochastic equation, $\mathbf{s_t}$ evolving as:

$$\mathbf{s_t} = A\mathbf{s_{t-1}} + B\mathbf{w_t}, \ \textit{where}\ A = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -0.8 & 0 & 1.8 & 0 \\ 0 & -0.8 & 0 & 1.8 \end{pmatrix}$$

Matrix A represents the deterministic evolution of the particle, $\mathbf{w_t}$ is a vector of standard normal variates and BB^T is the process noise covariance. The coefficients of A have been set empirically for now, describing an exponentially decreasing speed. The determination of B is achieved by noting that probabilities of random motion are isotropic: the correponding probability density function is described in Eq. 1, where i, j, k represent the three directions on the hexagonal grid (see fig. 2d). Fitting the hexagonal grid on the rectangular array parametrized by $x = i$ and $y = k$ coordinates, the probability density function can be expressed as Eq. 2.

$$p(i, j, k) \propto \exp(-(i^2 + j^2 + k^2)) \quad (1)$$
$$p(x, y) \propto \exp(-(x^2 + y^2 + (x + y)^2)) \quad (2)$$

which has a covariance matrix proportional to : $BB^T = \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix}$

Particle dynamics are therefore almost defined for a single panel. The scale factor for the covariance matrix is however unknown and is determined empirically.

Measurements. At each time step, the different cell values are extracted using the instrument model. For each particle, the 3D vector $\mathbf{v}$ that corresponds to the 2D cell it occupies is extracted from the precomputed correspondence table. The segments that correspond to the projection in the 2D image of the borders of a cylinder of direction vector $\mathbf{v}$ passing through the insertion point $\mathbf{T}$ are then extracted according to the instrument model previously described.

The image measurements we use are based on contour detection: image derivatives along the $\mathbf{x}$ and $\mathbf{y}$ axis are approximated by convoluting the image with corresponding Sobel kernels. The two components being extracted, the gradient vector $\mathbf{G_i}$ is available in each image point. We define a score, calculated for each segment : $\sigma_j = \sum_{i=1}^{m_j}(\mathbf{G_i} \cdot \mathbf{n_j})$, where $j = (1, 2)$, $\mathbf{n_j}$ is the unit vector normal to the segment j and m_j is the number of points in segment j. Finally the score $\pi = |\sigma_1 - \sigma_2|$ is given to the cell and the particle (n_1 and n_2 have opposite directions). The instrument axis is then computed as the average of all the particle vectors weighted by their score.

2.6 Determination of the Instrument Depth

Once the instrument axis has been computed, we determine the tip position along this axis by looking for a transition point between the instrument and the background. We developed an algorithm using a sliding window moving along the projection of the instrument's axis in the 2D image. For each position of the sliding window, we find the optimal threshold value dividing the pixels into two distinct classes according to Otsu's method [5]. We consider that the transition between the intrument's body and its tip is the point where the inter-class variance will be maximal. Therefore we store the position of the sliding window for which the inter-class variance is maximal,the position of the middle of this window returning the position of the tip of the instrument.

3 Results

Our first evaluation parameter is the angular error between the theorical instrument's axis and the computed one, in 3D and in the projected image. First, we tested the validity of our method with simulated data. We built a fictive sequence using an abdominal background image on which we pasted a black mask, corresponding to known 3D positions and orientations of the instrument. Its dynamic law was governed by the same equation as the particles (addition of a deterministic and a stochastic evolution).

We then tested our method on a testbench modeling a laparoscopic surgery, and compared our results with 1) manual segmentation of the tool in the image followed by 3D reconstruction in the same fashion as in 2.4 and 2) 3D results obtained using an Optical localizer. Our testbench consisted of a fixed endoscope imaging a surgical instrument inserted through a 5mm trocart, with a background image of an abdominal cavity containing another tool (the blue tool in fig. 4). The angular error for each setup according to the number of particles and geode resolution can be found fig. 3. Fig 4 shows typical image results that we

Measure type	Simulated data		Testbench data			
			Manual segmentation		Optical localizer	
	Mean	Stdv	Mean	Stdv	Mean	Stdv
Best particle (3D)	**9.2°**	**3.5°**	**30.4°**	**14.7°**	**14.8°**	**8.20°**
Particles average (3D)	**8.3°**	**5.0°**	**34.0°**	**11.9°**	**8.6°**	**5.8°**
Best particle (2D)	**2.6°**	**1.7°**	**5.7°**	**3.8°**	**1.76°**	**1.47°**
Particles average (2D)	**2.5°**	**1.6°**	**6.6°**	**5.3°**	**1.77°**	**1.46°**

Fig. 3. Results for 1050 consecutive frames, using 2000 particles and a geode resolution of 163842 cells covering the 3D sphere

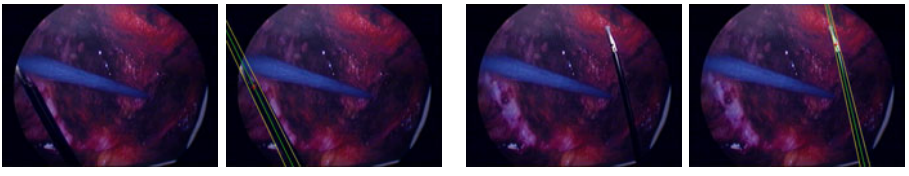


Fig. 4. Examples of the 2D results of an instrument's borders and tip detection in laparoscopic images

obtained on our testbench. Working with 768x576 images acquired by a clinical endoscope (no downsampling was performed), a geode resolution of 163842 cells and 500 or 2000 particles, our algorithm computes at the speed of respectively 16 and 8.6Hz with an Intel Xeon 2.67GHz, 3.48Go RAM PC.

Our second evaluation parameter was the 2D tip position in the image, which was compared to the tip position manually clicked on the images, with the error expressed in pixels. Our mean error is evaluated at 27.8 pixels for the same data set as in fig. 3.

4 Discussion - Conclusion

Our tracking system allows a surgeon to track the motion of his tools during an intervention, in order to either automatically control the movements of the robot, or to evaluate the quality of the surgeon's gesture. As shown by the results, our precision in the 2D orientation of the instrument inside the image is precise, robust and quick enough to allow the first anticipated use. The precision in the 3D orientation and the depth evaluation are still imperfect and should call for several improvements in our algorithm.

Increasing the number of cells in the geode (i.e. its resolution) does not slow down the algorithm as the correspondence table between the cells and 3D orientations is computed only once. However a limit is reached when the order is superior to 128 (163842 cells), as the surface of a cell at a distance of 35cm (the instrument's length) from the insertion point will be 3,13mm^2, which is approximatively the surface of the instrument tip.

Simulated data highlights the intrinsic error of the method, linked to the statistical and dynamical model, as the insertion point remains fixed, when testbench data shows as well the error due to the motion of the insertion point.

We now plan several improvements of our method to improve the 3D orientation detection and ensure its compliance with clinical conditions: first, we plan to allow the insertion point to move freely around its computed position , as it is the case in real surgery (in an amplitude range described in [10]), following a white noise or an evolution law which needs to be determined. Secondly, we plan to detect automatically the insertion point at the beginning of the procedure by extracting lines using the Hough transformation and the same pivot procedure as presented in section 2.4 (time-requesting computation, not adapted to real-time tracking).

Our next step will be to test our method on anatomic specimens mimicking true surgical interventions, in order to improve the deterministic and stochastic matrices describing the evolution of the particles - currently empirically evaluated - and to study their influence on the results. Improvements will also be brought to the tip detection in order to include an evolution model similar to that used for the axis orientation, thus smoothing aberrant measurements which pollute our mean result.

References

1. Aggarwal, R., Grantcharov, T., Moorthy, K., Milland, T., Darzi, A.: Toward feasible, valid, and reliable video-based assessments of technical surgical skills in the operating room. Annals of Surgery 247(2), 372 (2008)
2. Cano, A.M., Gayá, F., Lamata, P., Sánchez-González, P., Gómez, E.J.: Laparoscopic tool tracking method for augmented reality surgical applications. In: Bello, F., Edwards, E. (eds.) ISBMS 2008. LNCS, vol. 5104, pp. 191–196. Springer, Heidelberg (2008)
3. 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 11(5-6), 429–442 (2005)
4. Isard, M., Blake, A.: Condensation—conditional density propagation for visual tracking. International Journal of Computer Vision 29(1), 5–28 (1998)
5. Otsu, N.: A threshold selection method from gray-level histograms. IEEE Transactions on Systems, Man and Cybernetics 9(1), 62–66 (1979)
6. Randall, D.A., Ringler, T.D., Heikes, R.P., Jones, P., Baumgardner, J.: Climate modeling with spherical geodesic grids. Computing in Science and Engineering 4, 32–41 (2002)
7. Sherman, V., Feldman, L.S., Stanbridge, D., Kazmi, R., Fried, G.M.: Assessing the learning curve for the acquisition of laparoscopic skills on a virtual reality simulator. Surgical Endoscopy 19(5), 678–682 (2005)
8. Touijer, K., Guillonneau, B.: Laparoscopic radical prostatectomy: A critical analysis of surgical quality. European Urology 49(4), 625 (2006)
9. Uecker, D.R., Lee, C., Wang, Y.F., Wang, Y.: Automated instrument tracking in robotically assisted laparoscopic surgery. Journal of Image Guided Surgery 1(6), 308–325 (1995)
10. 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 Robotics Research 26(11-12), 1173–1190 (2007)
11. Zhang, Z.Y.: A flexible new technique for camera calibration. IEEE Transactions On Pattern Analysis and Machine Intelligence 22(11), 1330–1334 (2000)

Tracking and Characterization of Fragments in a Beating Heart Using 3D Ultrasound for Interventional Guidance

Paul Thienphrapa[1,2], Haytham Elhawary[1], Bharat Ramachandran[1], Douglas Stanton[1], and Aleksandra Popovic[1]

[1] Philips Research North America, Briarcliff Manor, NY, USA
aleksandra.popovic@philips.com
[2] ERC CISST/LCSR, Johns Hopkins University, Baltimore, MD, USA

Abstract. Fragments generated by explosions and similar incidents can become trapped in a patient's heart chambers, potentially causing disruption of cardiac function. The conventional approach to removing such foreign bodies is through open heart surgery, which comes with high perioperative risk and long recovery times. We thus advocate a minimally invasive surgical approach through the use of 3D transesophageal echocardiography (TEE) and a flexible robotic end effector. In a phantom study, we use 3D TEE to track a foreign body in a beating heart, and propose a modified normalized cross-correlation method for improved accuracy and robustness of the tracking, with mean RMS errors of 2.3 mm. Motion analysis of the foreign body trajectory indicates very high speeds and accelerations, which render unfeasible a robotic retrieval method based on following the tracked trajectory. Instead, a probability map of the locus of the foreign body shows that the fragment tends to occupy only a small sub-volume of the ventricle, suggesting a retrieval strategy based on moving the robot end effector to the position with the highest spatial probability in order to maximize the possibility of capture.

1 Introduction

Penetration of a fragment into the heart is a common injury in both civilian accidents and military warfare [9], [11]. Small caliber bullets and small shell fragments with low velocity tend to circulate freely in the chambers, potentially leading to arrhythmia, occlusion, and possibly death [3], [8]. The most frequent treatment approach to removing these fragments involves open surgery via a median sternotomy and cardiopulmonary bypasss (CPB) [9], a highly invasive procedure requiring a long recovery period. Risks of this procedure include bacterial mediastinitis, hemolysis, blood clotting, and air embolism.

The long term objective of this work is to develop a flexible robotic system for minimally invasive retrieval of fragments from a beating heart. In the envisioned clinical scenario, the end effector is inserted transapically into the heart (Fig. 1 [*left*]) after selection of the fragment via preoperative imaging. Then, under ultrasound guidance, the robot moves to capture the fragment; distal dexterity is

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 211–218, 2011.

needed to adequately cover the workspace. A minimally invasive approach can significantly improve the management of cardiac fragments by reducing risk and mortality, improving postoperative recovery, and potentially reducing operating room times. The objectives of the present study are (1) to demonstrate the utility of 3D transesophageal echocardiography (TEE) in tracking a target moving rigorously against a dynamic background; (2) to develop real-time tracking of the fragments; and (3) to analyze its motion in the context of a robotic system.

The use of 3D ultrasound for interventional systems is a fairly recent development due to high cost, low resolution, and inability to access real-time data streams [2]. In [7], markers were attached to an instrument to allow for image-based tracking. Variability in the readings was less than 1 mm. The instrument was subsequently driven towards targets in a water tank, reaching them with an error of 1.2 mm [4]. Experiments with low-resolution 3D catheter transducers [10] yielded errors of around 3 mm. Targets were static in these experiments.

Ferrous shrapnel can be difficult to detect in ultrasound when adjacent to other hyperechoic surfaces. In [5], a variable magnetic field is used to induce vibrations in shrapnel, thus illuminating it under color flow Doppler and allowing it to be located within an error of 1.06 mm.

The performance requirements for tracking a foreign body in the heart are more akin to the 1-degree-of-freedom (DOF) heartbeat compensation device of [12]. Predictive control was employed to track heart structures in one dimension with positive results. On the other hand, cardiac motions are more predictable than that of an object moving freely within the heart. Due to a variety of influences, foreign bodies are prone to seemingly arbitrary motions. A predictive control scheme is less likely to be effective in this case. The main contribution of this paper is the tracking of an unpredictably moving target using 3D TEE, for the purpose of developing a retrieval strategy by a flexible robotic device inserted through the apex of the heart.

2 Experimental Setup

A system for studying the problem of minimally invasive evacuation of fragments from a beating heart is illustrated in Fig. 1 (*center*). It includes an ultrasound system, a beating heart phantom, and a workstation (2.3 GHz Intel Xeon, 4 GB of RAM) that acquires live streaming ultrasound volumes over TCP/IP.

The Philips iE33 xMATRIX Echocardiography System is used with the X7-2t 3D TEE probe. Each image has a resolution of 128×48×112 voxels of size 0.81, 0.96, and 0.98 mm, spanning a field of view of 60° azimuth, 30° elevation, and 12 cm depth respectively. Gain and compression are set at 47% and 40 dB.

The beating heart phantom is a custom developed multi-modality phantom compatible with X-ray, ultrasound, and MR imaging. Made of polyvinyl acetate (PVA), it is a full replica of a human heart and resides in an acrylic glass water tank. Two servo-actuated pneumatic pistons pump water into and out of the phantom to create the deformable effect of a heartbeat and blood flow. Each piston pumps about 18 ml of water per heartbeat; its motion is given in Fig. 1

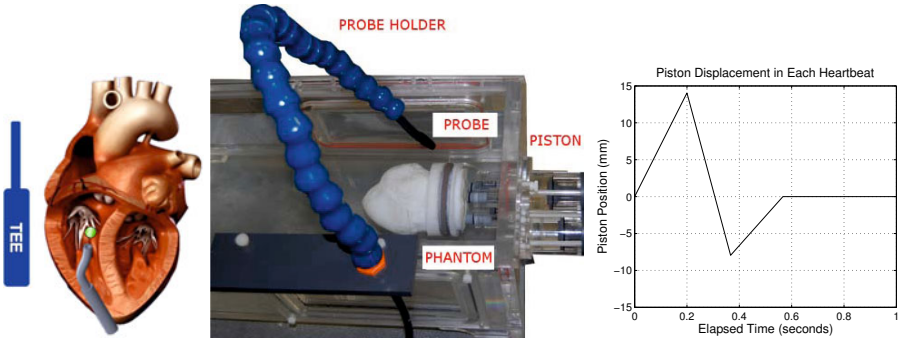


Fig. 1. (*Left*) Robotic fragment retrieval from the heart under 3D TEE guidance. (*Center*) Arrangement of the TEE probe and the heart phantom. (*Right*) Piston displacement in the beating heart phantom for one heartbeat.

(*right*). An experimental heart rate of 1 Hz and stroke volume lower than in healthy humans were chosen to mimic surgical [6] and post-injury cardiovascular conditions.

As an early approximation, a 3.2 mm steel ball was selected to act as a fragment due to its likeness to the clinical case in terms of size (typically 2-5 mm) and material. The fragment was imaged under 3D TEE at 20 frames per second (fps) and manually segmentated from five datasets (n=5) of 400 frames (20 seconds) each. This duration was empirically found to capture most of the activity that would occur in a given trial. Furthermore, 20 heart cycles permit robust testing of online tracking methods—a previous study reports on an *in vivo* instrument tracking experiment using a five-second window [4]. Fig. 2 shows the fragment in the heart phantom manually outlined in the 3D TEE image.

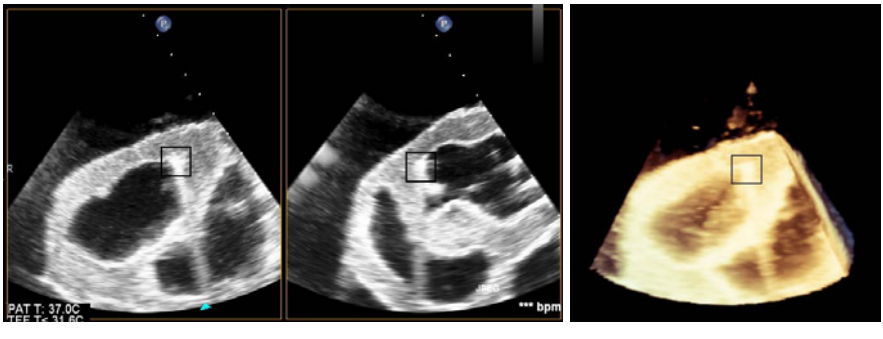


Fig. 2. (*Left and center*) Orthogonal slices showing the fragment (outlined) in the right ventricle of the heart phantom at rest. (*Right*) A three-dimensional view.

3 Real-Time Tracking of the Fragment

3.1 Approach

Normalized cross-correlation (NCC) is a well-known image processing algorithm used for template matching. The correlation coefficient at a voxel, $R(u, v, w)$, for a template $t(x, y, z)$ is given in (1). Normalization, accomplished by subtracting the mean from both the template and the image under it ($\overline{t}$ and $\overline{f}$ respectively) and dividing by their standard deviations, provides for robustness against changes in intensity, which can occur as both the fragment and background move.

$$R(u,v,w) = \frac{\sum_{x,y,z}\left[f(x,y,z) - \overline{f}\right]\left[t(x-u,y-v,z-w) - \overline{t}\right]}{\sqrt{\sum_{x,y,z}\left[f(x,y,z) - \overline{f}\right]^2 \sum_{x,y,z}\left[t(x-u,y-v,z-w) - \overline{t}\right]^2}} \quad (1)$$

The fragment was selected interactively from the first image frame to define the template for the algorithm to track in subsequent frames. A template of size 10×10×10 voxels was suitable for enclosing the entire fragment, while a search space of 40×30×30 was sufficient to capture displacements between frames. The low computation time of NCC (38 ms per frame) makes it especially suitable for tracking when real-time performance is required.

The coefficient R in (1) is a real number in $[-1, 1]$, with higher values indicating a close match and values tending toward -1 signifying a match of opposite intensity. The location of the maximum R (R_{max}) found in a search space is taken to be the position of the fragment. However, a weakness of NCC is that it relies on consistency in the apperance of the target, whereas apperances may

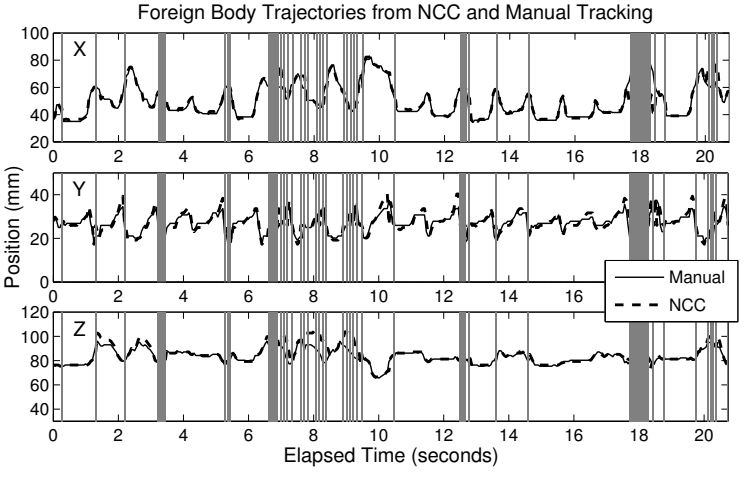


Fig. 3. Fragment trajectories based on NCC vs. manual tracking. Dropped frames are marked by vertical lines.

vary in ultrasound due to noise, artifacts, and low resolution. This can lead to the target being incorrectly identified in some frames. A low R_{max} suggests that NCC was not confident in its decision, and errant tracking can be avoided by ignoring such frames. In our experiments, a frame i is dropped if $R_{i,max}$ is less than c standard deviations below the mean of the dataset: $R_{i,max} < R_{mean} - c \cdot R_{sd}$. The value of c can be tuned to the level of confidence desired; we found c=0.9 to exhibit a reasonable compromise between spatial and temporal accuracy.

3.2 Results

Accuracy. Fig. 3 shows a trace of the fragment as determined by NCC, with manually-segmented fragment localization serving as a basis for evaluation.[1] Table 1 lists the number of dropped frames and accuracy for all datasets. NCC-based fragment tracking with detection of low confidence intervals (post-drop error) was accurate to within 2.3 mm rms after dropping 18.5% of frames. As expected, in the case of NCC without frame dropping (pre-drop error), tracking was lost in some frames resulting in mean errors that were almost twice as high (4.3 mm). For frames in which the reported fragment location is unavailable or doubtful, it is preferable to have the system wait until tracking is recovered, allowing the robotic retrieval instrument to stay in its current position or to be placed at a safe location if the uncertainty is prolonged over several frames.

Table 1. Accuracy of NCC referenced to data extracted manually

	Datasets					Aggregate			
Parameter	1	2	3	4	5	Min	Mean	Max	SD
Pre-Drop Error (mm rms)	6.2	4.8	5.3	2.5	2.6	2.5	4.3	6.2	1.7
Post-Drop Error (mm rms)	2.0	3.1	2.6	1.8	2.1	1.8	2.3	3.1	0.5
% Frames Dropped	20.9	15.5	20.1	22.9	13.3	13.3	18.5	22.9	4.0
$100 \times R_{mean}$	80.4	71.0	83.6	85.1	84.0	71.0	80.8	85.1	5.8
$100 \times R_{sd}$	9.2	8.5	10.3	4.8	3.7	3.7	7.3	10.3	2.9

Table 2. Fragment motion parameters obtained from 3D TEE images

	Datasets					Aggregate			
Parameter	1	2	3	4	5	Min	Mean	Max	SD
Range (mm)	51.4	49.0	47.5	23.5	29.8	23.5	40.2	51.4	12.7
Speed (mm/s)	336.2	334.2	530.0	208.7	308.3	208.7	343.5	530.0	116.5
Accel. (m/s^2)	5.3	9.5	9.1	6.2	8.7	5.3	7.8	9.5	1.9

Motion Characterization. Table 2 summarizes the extracted motion parameters. The average range of motion is 40.2 mm over three axes. While there is some correspondence between fragment motion and heartbeats (see Fig. 3), there is an element of motion that can seem arbitrary. Frequency domain analysis (Fig. 4) supports this observation with the primary peak located at 1 Hz,

[1] Video posted at http://www.youtube.com/watch?v=u-KYpbS33-8.

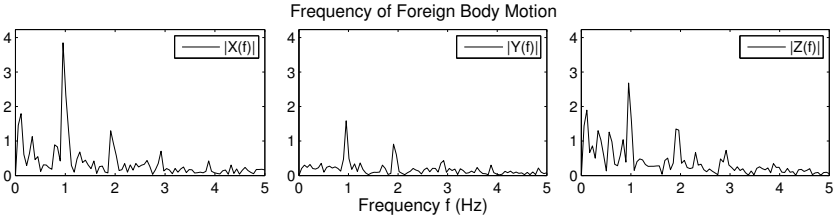


Fig. 4. Frequency components of fragment motion

but with spectral power at both lower (drift) and higher frequencies (heartbeat harmonics) suggesting a more complex behavior. Previous studies have found higher frequency components in the mitral valve as well [12].

4 Probability Map

The fragment reaches speeds of about 343.5 mm/s. Similar speeds have been reported for the heart wall (300 mm/s [1]) and mitral valve (200 mm/s [12]). Acceleration (7.8 m/s^2 on average) is higher than figures reported for the mitral valve (3.8 m/s^2 [12]). A retrieval strategy based on following the trajectory of the foreign body using a robotic end effector would be extremely challenging given these high speeds and accelerations. Furthermore, the uncertainties inherent in a beating heart environment may make it hazardous to do so. We thus propose a different approach for retrieval of the foreign body, wherein the real-time tracking method described in this paper is used to build a spatial probability map of the fragment, based on its past locations. The map can then be used to position a dexterous robotic device in the most visited locations of the fragment in order to maximize the possibility of capture.

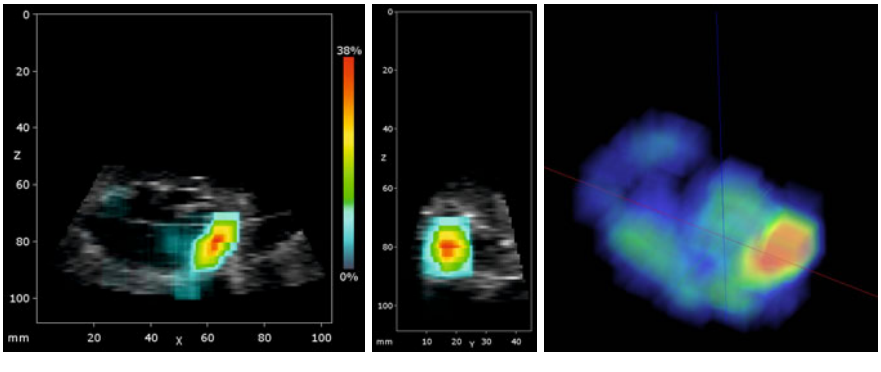


Fig. 5. Probability map of the fragment position. (*Left and center*) Coronal and sagittal slices of the map. (*Right*) Three-dimensional view.

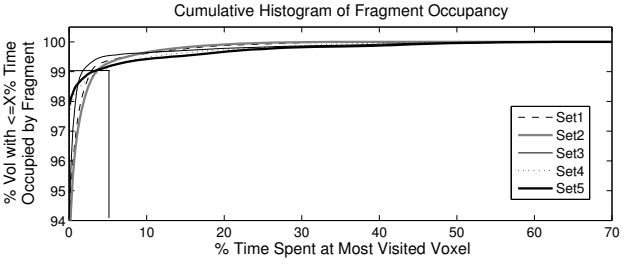


Fig. 6. Cumulative histograms showing how frequently voxels are visited by the fragment. Around Y=94% of the volume is never touched (X=0%), and roughly Y=99% of the volume is occupied at most X=5% of the time.

Table 3. Tendency of fragment to visit certain locations. The first row describes the percentage of the volume occupied at least 25% of the time. The second lists the maximum percentage of time that any one voxel was occupied.

	Datasets					Aggregate			
Parameter	1	2	3	4	5	Min	Mean	Max	SD
25% FB Occupancy	0.07	0.03	0.17	0.19	0.24	0.03	0.14	0.24	0.09
Max Occupancy Level	39.8	38.0	57.5	54.0	63.0	38.0	50.5	63.0	11.1

A qualitative representation of the fragment motion, generated using the tracking approach described above, is shown in Fig. 5. Warmer (red) regions indicate where the fragment is more likely to be located based on fragment traces. The observed tendency of the fragment to be localized in a subsection of the overall ventrical volume suggests the viability of a more relaxed retrieval approach based on aiming the dexterous robotic end effector at the volume of highest probability of fragment location. Histograms (Fig. 6 and Table 3) show that on the average across datasets, the fragment spends up to 50.5% of the time in the most visited voxel. Although it is shown in the frequency domain (Fig. 4) that the motion of the fragment is more irregular than the motion of the heart wall, predictive models using the spatial position of the fragment can be used to move the robot to a 'waiting' position, i.e. a position of high probability.

5 Conclusions and Future Work

In this paper, the motion of a fragment in a beating heart phantom is quantified using motion tracking in 3D TEE images. We show that the fragment motion is fast, abrupt, and often unpredictable. This provides valuable insight regarding the design of an interventional system that is capable of catching the particle. The next milestones will involve the design of a robot control scheme that uses the spatial probability map to safely retrieve the fragment, as well as studies of different fragment counts and physical properties.

References

1. Kostis, J.B., Mavrogeorgis, E., Slater, A., Bellet, S.: Use of a range-gated, pulsed ultrasonic Doppler technique for continuous measurement of velocity of the posterior heart wall. Chest 62(5), 597 (1972)
2. Krupa, A., Fichtinger, G., Hager, G.D.: Real-time motion stabilization with B-mode ultrasound using image speckle information and visual servoing. Int. J. Robot. Res. 28(10), 1334–1354 (2009)
3. Marshall, A.J., Ring, N.J., Newman, P.L.: An unexplained foreign body in the myocardium. J. Royal Soc. Med. 95, 250–251 (2002)
4. Novotny, P.M., Stoll, J.A., Vasilyev, N.V., del Nido, P.J., Dupont, P.E., Zickler, T.E., Howe, R.D.: GPU-based real-time instrument tracking with three-dimensional ultrasound. Med. Image Anal. 11, 458–464 (2007)
5. Rogers, A.J., Light, E.D., Smith, S.W.: 3-D ultrasound guidance of autonomous robot for location of ferrous shrapnel. IEEE Trans. Ultrason., Ferroelectr., Freq. Control 56(7), 1301–1303 (2009)
6. Ryan, K.L., Cooke, W.H., Rickards, C.A., Lurie, K.G., Convertino, V.A.: Breathing through an inspiratory threshold device improves stroke volume during central hypovolemia in humans. J. Appl. Physiol. 104(5), 1402–1409 (2008)
7. Stoll, J.A., Novotny, P.M., Howe, R.D., Dupont, P.E.: Real-time 3D ultrasound-based servoing of a surgical instrument. In: IEEE Int. Conf. on Robotics and Automation, pp. 613–618 (2006)
8. Symbas, P.N., Picone, A.L., Hatcher, C.R., Vlasis-Hale, S.E.: Cardiac missiles. A review of the literature and personal experience. Ann. Surg. 211(5), 639–647 (1990)
9. United States Department of Defense: Emergency War Surgery: Third United States Revision, s.l., 16: Thoracic Injuries (2004)
10. Whitman, J., Fronheiser, M.P., Smith, S.W.: 3-D ultrasound guidance of surgical robotics using catheter transducers: feasibility study. IEEE Trans. Ultrason., Ferroelectr., Freq. Control 55(5), 1143–1145 (2008)
11. Williams, J.C., Elkington, W.C.: Slow progressing cardiac complications—a case report. J. Chiropr. Med. 7, 28–33 (2008)
12. Yuen, S.G., Kettler, D.T., Novotny, P.M., Plowes, R.D., Howe, R.D.: Robotic motion compensation for beating heart intracardiac surgery. Int. J. Robot. Res. 28(10), 1355–1372 (2009)

Model-Based Fusion of Multi-modal Volumetric Images: Application to Transcatheter Valve Procedures

Saša Grbić[1,3], Razvan Ionasec[1], Yang Wang[1], Tommaso Mansi[1], Bogdan Georgescu[1], Matthias John[2], Jan Boese[2], Yefeng Zheng[1], Nassir Navab[3], and Dorin Comaniciu[1]

[1] Image Analytics and Bioinformatics, Siemens Corporate Research, Princeton, USA
[2] Siemens AG, Healthcare Sector, Forchheim, Germany
[3] Computer Aided Medical Procedures, TU Munich, Germany

Abstract. Minimal invasive procedures such as transcatheter valve interventions are substituting conventional surgical techniques. Thus, novel operating rooms have been designed to augment traditional surgical equipment with advanced imaging systems to guide the procedures. We propose a novel method to fuse pre-operative and intra-operative information by jointly estimating anatomical models from multiple image modalities. Thereby high-quality patient-specific models are integrated into the imaging environment of operating rooms to guide cardiac interventions. Robust and fast machine learning techniques are utilized to guide the estimation process. Our method integrates both the redundant and complementary multimodal information to achieve a comprehensive modeling and simultaneously reduce the estimation uncertainty. Experiments performed on 28 patients with pairs of multimodal volumetric data are used to demonstrate high quality intra-operative patient-specific modeling of the aortic valve with a precision of 1.09mm in TEE and 1.73mm in 3D C-arm CT. Within a processing time of 10 seconds we additionally obtain model sensitive mapping between the pre- and intra-operative images.

1 Introduction

There has been a major trend in cardiac therapy towards minimally invasive transcatheter procedures to reduce the side effects of classical surgical techniques. Instead of full sternotomy, instruments and devices are introduced through small incisions, advanced through vessels and positioned to perform various procedures [1]. Without direct access and view to the affected structures those interventions are usually performed in so-called Hybrid ORs, operating rooms outfitted with advanced imaging equipment. Thus, procedures such as the Transcatheter Aortic Valve Replacement (TAV) are permanently guided via real-time intra-operative images provided by C-arm X-ray and Transesophageal Echocardiography systems [2].

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 219–226, 2011.

Traditionally the field of medical image analysis has been focusing on the construction of patient-specific anatomical models from well established diagnostic modalities (e.g. CT and MR) to aid disease analysis and treatment planning [3]. In the context of valvular disease management, the authors in [4] and [5] proposed the modeling of the aortic valve from cardiac CT. Models of the mitral valve from MR have been proposed in [6]. However, these methods have not been developed to cope with the reduced quality and contrast characteristic to intra-operative images, hence their usage is strictly limited to pre-operative decision making.

Delivering high-quality models into the operating room to guide cardiac therapy will be a major focus of future clinical applications. The fusion of pre-operative models with intra-operative images for mitral valve replacement has been proposed by [7]. The pre-op model of the mitral valve was annotated from CT data and registered into the intra-op MR and echocardiography images. Major limitation is the required tracking equipment and semi-automatic delineation of the mitral annuls. Alternatively, multi-modal image registration has been proposed to fuse multi-modal data. In [8] the mutual information is used as the metric to cope with the intensity inconsistencies between CT and MR. In [9] an atlas-based approach was presented to track the myocardium and left and right ventricles from MR data. The registration is used to align the cardiac atlas to the patient data. However, these methods are computationally expensive, and without the appropriate guidance of a shape prior likely to converge into local minima.

As we seek to provide both, the fusion of pre- and intra-operative imaging and patient-specific models of relevant anatomical structures, the field of simultaneous registration and segmentation is important to our work. In [10] a probabilistic framework was proposed where registration is performed jointly with segmentation. It was applied to the segmentation of brain tissues and their substructures in uni-modal MR data. [11] recently proposed a method to jointly segment the prostate and provide registration in MR. It introduced point sets to allow fast initial registration. However most of these methods suffer from long run-times, as the problem of registration and segmentation is trying to be solved simultaneously, and the necessity of manual initialization.

We propose a novel method to fuse pre-operative and intra-operative information by jointly estimating anatomical models from multiple image modalities. Thereby high-quality patient-specific models are integrated into the imaging environment of operating rooms to guide cardiac interventions. Robustness and efficiency are achieved by relying on machine learning techniques to drive the joint estimation process whereby similarities between multiple modalities are exploited. Statistical models of the anatomy are utilized within the probabilistic estimation framework to ensure physiologically compliant results. The main benefits of our method are: 1) Completeness - by exploiting the complementary information from multiple modalities, 2) Robustness - by exploiting the redundant information from multiple modalities to reduce the estimation uncertainty, and 3) Fusion - by obtaining a model-sensitive integration of the pre-operative and intra-operative modalities.

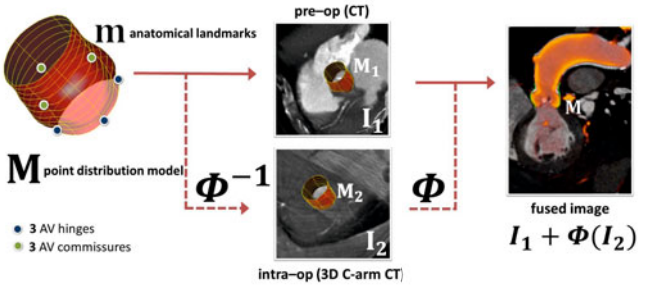


Fig. 1. Diagram of the problem formulation showing the surface model M, anatomical landmarks m, transformation ϕ to map the intra-op image I_2 to the pre-op data I_1

2 Problem Formulation

Our goal is to estimate a 3D anatomical patient-specific model M from volumetric multi-modal datasets I_1 and I_2, where I_1 is the pre-op and I_2 the intra-op image, and the transformation ϕ which maps the intra-op model M_2 to the pre-op model M_1 (see Fig 1).

$$(\hat{\phi}, \hat{M}) = \arg\max_{M,\phi} \log P(M, \phi \,|I_1, I_2\,) \tag{1}$$

ϕ is composed of an affine- A and a non-linear warping transformation D, $\phi = D\ A$. D is modeling the small deformation of M due to respiration and uncertainties in the acquisition phase between the pre- and intra-op data. The model M is represented as a point distribution model. Using the transformation ϕ the pre- M_1 and intra-op M_2 models can be computed: $M = M_1$, $M = D\ A\ M_2$ and $M_2 = A^{-1}\ D^{-1}\ M$.

3 Method

In general finding an optimal solution to Eqn. 1 is difficult and has high computational cost therefore we approximate the problem by expanding the formulation and exploiting independencies. In addition a shape constraint term is added to restrict the estimated model M in a shape space built from a database of annotations.

$$(\hat{\phi}, \hat{M}) = \arg\max_{M,\phi} \log \left(P(M\,|I_1\,) \cdot P(M\,|\phi(I_2)\,) \cdot P(M\,|I_1, \phi(I_2)\,) \cdot P(M, \phi\,|\mu, \Sigma\,)\right) \tag{2}$$

All the probabilities in our formulations are modeled using robust learning based algorithms. The first $P(M\,|I_1\,)$ and the second term $P(M\,|\phi(I_2))$ define the independent model estimations in the multi-modal images I_1 and I_2. As proposed in [4] a classifier is trained using the probabilistic boosting tree and Haar-features to estimate the posterior probability. The best model parameters for M are selected based on a joint probability term $P(M\,|I_1, \phi(I_2)\,)$ explained in chapter 3.1.

The transformation ϕ is modeled as a warping transform with Gaussian radial basis functions. The last term $P(M, \phi \,|\mu, \Sigma)$ symbolizes a regularization of the shape M and the transformation ϕ based on the learned statistical shape model defined as a Gaussian distribution with the mean μ and the covariance matrix Σ learned from manual annotations. Both the affine A and the non-linear transformation D are updated in this stage. A bias is applied towards the pre-op model $M = M_1$ as the model estimation is more robust in the pre-op images. In our case I_1 represents the CT image and I_2 the TEE and 3D C-arm CT image.

In this work we focus on the estimation of the aortic valve model. The valve is modeled hierarchically using two layers. On the coarse level it is represented as a landmark model m with 6 points (3 commissures and 3 hinges). They define the most important morphological and functional properties of the valve. The finer level is defined as a point distribution model M with 1440 points spread along a 36×20 parametric grid.

3.1 Similarity Learning

The joint term $P(M \,|I_1, \phi(I_2))$ in Eqn. 2 exploits the similarities between the models from the multi-modality images. Similarity functions proposed in the current literature, such as mutual information or cross correlation, could be used but as mentioned in [12] learning the similarity for a specific problem yields better performance.

We employ a boosting framework in order to train a cascade of strong classifiers. Each strong classifier $\boldsymbol{F}_{strong}$ consists of k weak classifiers $\boldsymbol{F}_{weak}$ which learn the similarity between pairs of image patches $I_{S1} \in I_1$ and $I_{S2} \in I_2$, $\boldsymbol{F}_{weak}(I_{S1}, I_{S2})$. The weak learners are constructed based on Haar-like features extracted locally from rectangular patches I_{S1} and I_{S2} from image slices sampled perpendicular to the tubular aortic root surfaces M_1 and M_2. The patch size is fixed for both modalities.

The weak learner is modeled as a 2D piecewise constant function defined on a 2D feature space by the feature responses of $h(I_{S1})$ and $h(I_{S1})$. The 2D feature space is separated in equal rectangular non-overlapping regions. Therefore we quantize the feature responses from both modalities in 64×64 bins whereby the values are scaled between the minimum and maximum feature responses $h(I_{S1})$ and $h(I_{S1})$.

$$\boldsymbol{F}_{weak}(I_{S1}, I_{S2}) = \sum_{b=1}^{B} \sum_{c=1}^{C} \beta_{b,c} \, R_{b,c} \, [h(I_{S1}) \times h(I_{S2})] \tag{3}$$

where B and C are the bin numbers for the feature responses in each modality and $\beta_{b,c}$ symbolizes the constant associated with the region $R_{b,c}$ representing a bin in the 2D feature space. As in [13] the optimal weights $\beta_{b,c}$ would be determined by fitting a least-squares regression function. During detection a probability for each weak classifier is evaluated by extracting Haar-features from pairs of image patches. The features are assigned to a bin $R_{b,c}$ based on the feature response and multiplied with the corresponding weight $\beta_{b,c}$.

A cascade of l strong classifiers $\boldsymbol{F}_{strong}$ is trained in order to determine the posterior probability $P(M\,|I_1, \phi(I_2)\,) = \boldsymbol{S}(I_{S1}, I_{S2})$ of the similarity function.

3.2 Model-Based Fusion Approach

The first stage in our hierarchical model estimation algorithm consists of pre-aligning the multi-modal images using the anatomical landmarks. The affine transformation A is estimated by obtaining a least-squares solution based on the independently detected landmarks m_1 from the image I_1 and m_2 from the image I_2. The landmark detectors are trained using the probabilistic boosting tree classifier and Haar-like features. The surface M is initialized by learning a correlation model between measurements extracted from the landmarks m_1 and the point distribution model M, as described in [14]. The nonlinear warping transformation D is set to identity. Based on A the model M can be projected to the image I_2.

In the optimization phase we apply an iterative approach. We sample candidates N_1 and N_2 along the surfaces normals of M_1 and M_2, and evaluate the probability $P(M\,|I_1\,)$ for each candidate $n_1 \in N_1$ and $P(M\,|\phi(I_2)\,)$ for each point $n_2 \in N_2$ using the learned detectors. The joint probability $P(M\,|I_1, \phi(I_2)\,)$ is determined by training a boosting classifier, as mentioned in chapter 3.1, to evaluate pairs of candidates. A cross product of the candidates $N_1 \times N_2$ is constructed and the highest probable candidate pair (n_i, n_j) is selected by multiplying the single modality probabilities with the joint term.

$$(n_i, n_j) = \arg\max_{n_i, n_j} \log\left(P(n_i\,|I_1\,) \cdot P(n_j\,|\phi(I_2)\,) \cdot P(n_i, n_j\,|I_1, \phi(I_2)\,)\right) \qquad (4)$$

The estimated candidate pairs are used to update the models M_1 and M_2. The second step of the iteration involves calculating the posterior probability $P(M, \phi\,|\mu, \Sigma\,)$ of M and ϕ based on the learned statistical shape models. This could be perceived as a regularization to the shape M. Thereby M_1 is projected to the PCA shape space using the largest 40 eigenvectors. ϕ is updated by computing the rigid transformation R based on the posterior probability of the pairs (n_i, n_j). D is updated by obtaining a least-squares solution to the warping transformation $\hat{D} = \arg\min \left\| T\ M_2 - D^{-1} M_1 \right\|^2$ using radial basis functions. Thereby

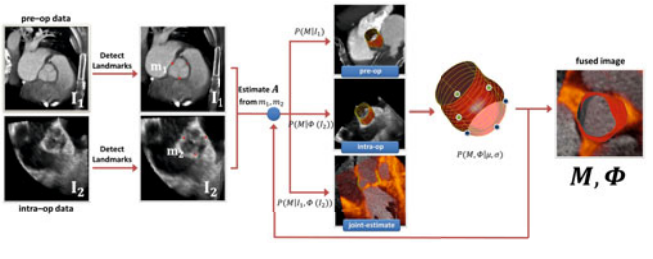


Fig. 2. Diagram showing the model based fusion approach for the estimation of the model M and the transformation ϕ

Fig. 3. Example of the joint aortic valve model estimation from pre- and intra-op volumetric data. The left 2 images show fused CT-TEE data sets and the right 2 images show fused CT-3D C-arm CT data. The mapping of the intra-op image I_2 to the pre-op image I_1 is done by the estimated non-linear transform ϕ.

the number of control points is much smaller than the number of shape points M. The algorithm converges in a small number of iterations. Figure 2 demonstrates the complete estimation approach.

4 Experimental Results

The most relevant intra-operative modalities with 3D capabilities (3D C-arm CT and TEE) in the OR environment were incorporated for evaluation. In total 56 volumes, 13 pairs of CT-TEE data sets and 15 pairs of CT-3D C-arm CT data pairs were selected to demonstrate the effectiveness of our method. This dataset was solely used for evaluation and not included in training. The ground-truth annotations were obtained from clinical experts by manually placing the anatomical landmarks in the pre- (m_1) and intra-op (m_2) images and finally delineating the aortic valve surfaces M_1 and M_2.

As our algorithm depends on the automatic detection of the anatomical landmarks m_1 and m_2 during the initialization step in order to estimate the affine transform A we evaluate their detection performance on the test dataset. For training 160 separate landmarks annotations in CT, 320 in TEE and 192 in 3D C-arm CT were used to train the landmark detectors. The error is computed as the Euclidian distance between the automatic estimation and the expert annotation. For the hinges we obtain an error of $2.40 \pm 0.81mm$ in CT, $2.56 \pm 0.71mm$ in TEE and $2.30 \pm 1.56mm$ in 3D C-arm CT and for the commissures $2.74 \pm 1.01mm$ in CT, $3.31 \pm 1.55mm$ in TEE and $2.98 \pm 1.44mm$ in 3D C-arm CT.

The mesh-to-mesh error was computed between the ground-truth annotations and the detected models in order to obtain quantitative results for the automatic surface estimation. Results shown in table 1 confirm that our model-based fusion estimation approach yields the best results.

In transcatheter aortic valve procedures both the selection of the appropriate stent size but also the positioning of it in the intra-op data has clinical significance. However in 3D C-arm CT the aortic valve annulus is not visible as the contrast is injected at the cusp area. Fusing the 3D C-arm image with pre-op CT data would allow the physician to properly examine the annulus area and enable accurate positioning of the stent during the procedure. We evaluate the error

Table 1. System precision for aortic valve surface model estimation in CT, TEE and 3D C-arm CT. Comparison between our novel model-based fusion approach and single modality estimations.

	single modality estimation			fusion approach		
	Mean	STD	Median	Mean	STD	Median
CT-TEE [mm]	1.22	0.23	1.13	1.09	0.22	1.10
CT-3D C-arm CT [mm]	1.96	0.54	1.99	1.73	0.49	1.79

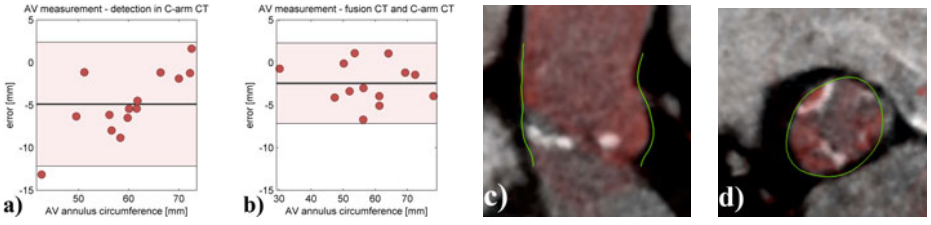


Fig. 4. Bland-Altman plots for the aortic valve annulus circumference measurement extracted from the model M with (a) independent detection in 3D C-arm CT and (b) fusion of pre-op CT and 3D C-arm CT. (c) and (d) are showing short and long axis views of the model M and the fused pre- and intra-op images $I_1 + \phi(I_2)$.

for the aortic valve annulus ring circumference, extracted from the estimated aortic valve model M, by comparing the result of the independent detection in 3D C-arm CT image and our model-based fusion approach. Quantitative and qualitative results are shown in figure 4.

5 Conclusion

In this paper, we propose a novel approach to estimate comprehensive patient specific models of the aortic valve by model-sensitive fusing of multimodal pre- and intra-operative data. Fast and robust machine learning techniques are employed during the estimation exploiting redundant and complementary information from the multimodal images. Thereby high-quality patient-specific models are integrated into the imaging environment of operating rooms to guide cardiac interventions. Comprehensive quantitative and qualitative experiments on the aortic valve modeling demonstrate the effectiveness of our approach with an accuracy of 1.09mm in CT-TEE and 1.73mm in CT-3D C-arm CT.

References

1. Walther, T., Chu, M., Mohr, F.: Transcatheter aortic valve implantation: time to expand? Curr. Opin. Cardiology 23, 111–116 (2008)
2. Vahanian, A., Alfieri, O., Al-Attar, N.: Transcatheter valve implantation for patients with aortic stenosis. Eur. J. Cardiothorac Surg. 34, 1–8 (2008)

3. Frangi, A.F., Rueckert, D., Duncan, J.S.: Three-dimensional cardiovascular image analysis. IEEE Transaction on Medical Imaging 21(9), 1005–1010 (2002)
4. Ionasec, R.I., Voigt, I., Georgescu, B., Wang, Y., Houle, H., 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 Transaction on Medical Imaging 29(9), 1636–1651 (2010)
5. Waechter, I., Kneser, R., Korosoglou, G., Peters, J., Bakker, N.H., Boomen, R.v.d., Weese, J.: Patient specific models for planning and guidance of minimally invasive aortic valve implantation. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6361, pp. 526–533. Springer, Heidelberg (2010)
6. Conti, C.A., Stevanella, M., Maffessanti, F., Trunfio, S., Votta, E., Roghi, A., Parodi, O., Caiani, E.G., Redaell, A.: Mitral valve modelling in ischemic patients: Finite element analysis from cardiac magnetic resonance imaging. In: Computing in Cardiology, pp. 1059–1062 (2010)
7. Linte, C.A., Wierzbicki, M., Moore, J., Little, S.H., Guiraudon, G.M., Peters, T.M.: Towards subject-specific models of the dynamic heart for image-guided mitral valve surgery. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part II. LNCS, vol. 4792, pp. 94–101. Springer, Heidelberg (2007)
8. Wells, W.M., Viola, P., Kikinis, R.: Multi-modal volume registration by maximization of mutual information. Medical Image Analysis 1(1), 35–51 (1996)
9. Lorenzo-Valdés, M., Sanchez-Ortiz, G.I., Mohiaddin, R.H., Rueckert, D.: Atlas-based segmentation and tracking of 3D cardiac MR images using non-rigid registration. In: Dohi, T., Kikinis, R. (eds.) MICCAI 2002. LNCS, vol. 2488, pp. 642–650. Springer, Heidelberg (2002)
10. Pohl, K.M., Fisher, J., Grimson, L., Kikinis, R., Wells, W.M.: A Bayesian model for joint segmentation and registration. NeuroImage 31(1), 228–239 (2006)
11. Gao, Y., Sandhu, R., Fichtinger, G., Tannenbaum, A.: A Coupled Global Registration and Segmentation Framework With Application to Magnetic Resonance Prostate Imagery. IEEE Transaction on Medical Imaging 29(10), 1781–1794 (2010)
12. Bronstein, Michael, A.M., Science, C., Israel, T., Michel, F., Paragios, N., Galen, E., Ile-de france, I.S.: Data Fusion through Cross-modality Metric Learning using Similarity-Sensitive Hashing. In: CVPR 2010, pp. 3594–3601 (2010)
13. Zhou, S., Shao, J., Georgescu, B., Comaniciu, D.: Boostmotion: Boosting a discriminative similarity function for motion estimation. In: CVPR 2006, pp. 1761–1768 (2006)
14. Grbić, S., Ionasec, R., Vitanovski, D., Voigt, I., Wang, Y., Georgescu, B., Navab, N., Comaniciu, D.: Complete valvular heart apparatus model from 4D cardiac CT. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6361, pp. 218–226. Springer, Heidelberg (2010), http://www.ncbi.nlm.nih.gov/pubmed/20879234

1D–3D Registration for Functional Nuclear Imaging

Christoph Vetter[1,2], Tobias Lasser[1], Thomas Wendler[1,3], and Nassir Navab[1]

[1] Chair for Computer Aided Medical Procedures (CAMP),
Technische Universität München, Germany
[2] Siemens Corporate Research, Princeton, USA
[3] SurgicEye GmbH, Munich, Germany

Abstract. Functional nuclear imaging systems like PET or SPECT provide unique information that is used extensively in diagnosis, but it has also proven very useful for image–guided interventions. In the case of SPECT and radio–guided surgeries, 1D gamma detectors called gamma probes are routinely used during interventions to localize hotspots in conjunction with pre–operative SPECT images, or more recently, intra–operative SPECT images. As the tissue is being manipulated during surgery, these SPECT images quickly lose their validity, necessitating either new scans, which is in most cases unfeasible, or requiring the surgeon to do a mental update of the available imagery. In this paper, we present a novel 1D–3D registration procedure for functional nuclear imaging that registers tracked intra–operative 1D probe readings to a pre– or intra–operatively acquired 3D functional image. This procedure allows incorporating prior knowledge during radio–guided surgeries, enabling rapid updates of the visualization in the case of tissue deformation without the overhead of an additional complete scan. We show results using phantom data as well as patient data.

1 Introduction

Functional nuclear imaging modalities like PET or SPECT provide specific information to localize in 3D even very small tumors in several oncological applications [2,3]. Translating this information to the operating room however faces several difficulties: the patient posture may change, treatment in between the imaging and surgery might affect the region of interest and finally, tissue manipulation during surgery invalidates the previous imaging information [1].

In the case of SPECT and radio–guided surgeries, 1D gamma detectors (also called gamma probes) are routinely used during interventions to localize hotspots in conjunction with pre–operative SPECT images, or more recently, intra–operative SPECT images [7] (see figure 1). The count rates from the gamma probe, presented visually as well as acoustically (like a Geiger counter), help the surgeon translating in his mind the available functional imaging information to the current situation.

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 227–234, 2011.

In this work we suggest a novel, automatic 1D–3D registration procedure to register 1D gamma probe readings, augmented by synchronized position and orientation data acquired by an optical tracking system, to pre– or intra–operative 3D SPECT imaging data, which has been invalidated due to tissue manipulation during surgery. We demonstrate our approach for the use case of sentinel lymph node biopsies (SLNB) in breast cancer. Sentinel lymph node biopsy is a technique to determine if breast cancer has spread into the lymphatic system in a low morbidity way and is now standard practice for early stage breast cancer patients [5]. In the SLNB procedure a low–energy radioactive tracer for SPECT is injected near the tumor, and the lymph nodes showing radioactivity as for example indicated by a hand–held gamma probe (i.e. the first nodes in the lymph drainage path of the tumor) are then excised and examined histologically for cancerous cells. Given the presence or absence of metastases in the excised nodes, different treatments are indicated making this technique of major importance in cancer staging. The possibility of selectively excising the first nodes in the drainage of the tumor with the help of this nuclear medicine approach allows better targeted surgery, less morbidity and shorter hospital stays for the patient. Technical advancements like 3D intra–operative SPECT imaging are poised to improve the accuracy of the SLNB procedure [6], but even there tissue manipulation during surgery poses difficulties that can be addressed using the introduced 1D–3D registration procedure in this work.

The technique suggested here is one step towards a future in which surgical instrument containing integrated detectors provide a constantly updated view of the radioactivity in tissue, allowing real–time updates of the guidance for the surgeon, which results in faster and more efficient surgery as well as a better quality control. This paper presents the methods employed for the proposed 1D–3D registration procedure and evaluates the procedure using a series of phantom experiments as well as patient data.

2 Methods

2.1 Input Data and Output

The registration method requires two inputs, a three–dimensional volume $V \subset \mathbb{R}^3$ (for example SPECT) and a series of tracked 1D probe measurements $b_j \in \mathbb{R}$, $j = 1, \ldots, k$ (for example from a gamma probe), which are acquired during the surgery. Here b_j denotes the count rate (counts per second) of the probe, and associated with each measurement b_j are the synchronized position and orientation $(p_j, o_j) \in \mathbb{R}^3 \times \mathbb{R}^3$ as acquired by the tracking device. The output of the method will be an updated 3D volume $\widehat{T}(V)$, where $\widehat{T} : \mathbb{R}^3 \to \mathbb{R}^3$ is an approximation of the optimal transformation transforming the no longer valid volume V into the desired volume $\widehat{T}(V)$.

The challenge in this registration method lies in the very limited number of probe measurements covering only a fraction of the volume of interest due to time and practical constraints. Several other factors also add complications,

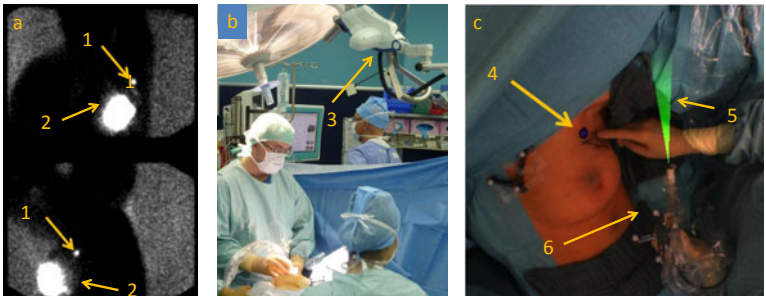


Fig. 1. a) shows two pre–operative projection angles of a scintigraphy for SLNB (top: frontal, bottom: left lateral) with the lymph node (1), and the injection site (2), b) shows the workstation with the gamma probe and the tracking device (3) during a SLNB procedure, c) shows an augmented reality view of the surgery area. The visual guidance overlay consists of the node in blue indicating the lymph node (4) and the triangle in green, indicating the field of view of the probe (5). The hand–held gamma probe is marked as 6.

namely the statistical nature of the detection process, the large field of view of the gamma probe as well as inaccuracies of the tracking system along with the lack of information of the tissue's attenuation and scattering properties.

In the particular case of SLNB, the probe measurements are typically acquired using gyrating movements around the suspected location of a lymph node, starting at the old location indicated by the SPECT, moving closer to the current location as guided by the probe measurements via visual or acoustic feedback. This means that large parts of the SPECT volume are not covered by probe measurements and that many measurement angles are missing.

To cope with this, the registration method partitions the SPECT volume into a fixed part V_f and a moving part V_m, $V_f \cup V_m = V$, $V_f \cap V_m = \emptyset$. This is accomplished by segmenting the lymph nodes $\mathcal{N} = \{N_1, \ldots, N_l\}$ in the SPECT volume, choosing the lymph nodes $\mathcal{M} \subset \mathcal{N}$ which are closest to the probe measurement positions p_j and then forming $V_m = \bigcup_{M \in \mathcal{M}} \mathrm{bbox}(M)$ as the union of spherical bounding boxes around the nodes in $\mathcal{M}$. We set $V_f = V \setminus V_m$. The actual registration procedure is then only applied to V_m, that is, the output of the method is $\widehat{T}(V_m)$, where $\widehat{T} : V_m \to \mathbb{R}^3$ reuses the previous notation of the unrestricted $\widehat{T}$.

2.2 Registration Procedure

For choosing a deformation model, it is important to note that the deformation model can only incorporate information that is provided by the nuclear probe. Even though the unlabeled tissue around the radioactively labeled nodes restricts the nodes' movement, no information about this tissue is available. In particular, surgical operations like cutting or cauterizing change tissue and tissue connectivity. Hence our deformation model treats each node in $\mathcal{M}$ as moving

independently. Furthermore, each node in $\mathcal{M}$ is assumed to not deform from its spherical shape. The deformation model then ends up as pure translations $\widehat{T}|_M$ for each $M \in \mathcal{M}$.

For the actual registration, the probe measurements b_j act as the *fixed image*, while the *moving image* are measurements simulated from the current transformed volume $T(V_m)$, where $T : V_m \rightarrow \mathbb{R}^3$ denotes the current estimate of $\widehat{T}$. The similarity measure employed is sum of squared differences (SSD). This results in the minimization of

$$\widehat{T} = \text{argmin}_T \sum_{j=1}^{k} \big(P_j\big(T(V_m)\big) - b_j\big)^2 + \text{reg}\big(T(V_m)\big), \tag{1}$$

where $P_j\big(T(V_m)\big)$ denotes the j–th simulated measurement from $T(V_m)$, and $\text{reg}\big(T(V_m)\big)$ denotes a regularization term. In this work we used $\text{reg}\big(T(V_m)\big) = \|T(V_m)\|_2^2$.

The resulting minimization problem can be solved with standard optimization methods. For this work we chose gradient descent, but other methods are applicable as well.

The most notable part in equation (1) are the mappings $P_j : V_m \rightarrow \mathbb{R}$, which perform the 1D–3D transformation step by simulating probe measurements at position/orientation (p_j, o_j) from the *moving image* $T(V_m)$. It can also be thought of as a forward projection in tomography terms. This requires a model of the detection process in the probe device, modeling characteristics like detector sensitivity, collimation and shielding. In this work we employ a simple solid angle physical model of the detection process, modeling the field of view, angular detection as well as distance sensitivity analytically [4]. Any source in $T(V_m)$ is modeled as a Poisson distribution dependent on the activity in $T(V_m)$, as is regularly done in emission tomography. In mathematical terms we discretize $T(V_m) = (x_i)_{i=1,\ldots,n}$ into n regular isotropic voxels, then

$$P_j\big(T(V_m)\big) = \sum_{i=1}^{n} model(p_j, o_j, x_i), \tag{2}$$

where $model(p_j, o_j, x_i)$ denotes the physical model, that is the contribution of voxel x_i to the measurement of the probe located at position/orientation (p_j, o_j).

3 Experiments and Results

3.1 Hardware Setup

SPECT/CT imaging was provided by a Symbia T6 system (Siemens, Healthcare, Germany), intra–operative SPECT (in short: ioSPECT) by the declipseSPECT cart system (SurgicEye, Germany). The gamma detector employed was a Gamma–Probe System (Crystal Photonics, Germany), and tracking was performed using a Polaris Vicra infrared optical tracking system (Northern Digital, Canada) with reflective markers attached to both the gamma probe and to the scanning target (phantom or patient).

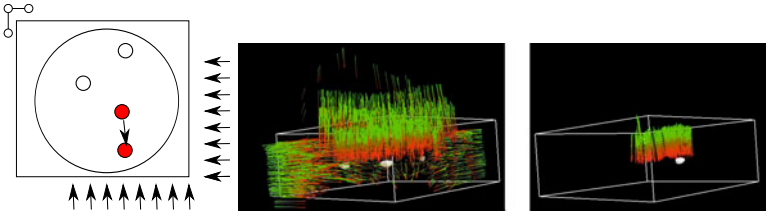


Fig. 2. *Left:* the phantom with three hollow spheres filled with a ^{99m}Tc solution. The arrows illustrate scanning directions for ioSPECT, scanning from top not shown. The red sphere is moved by 2cm to simulate deformation during surgery. *Middle:* full scan suited for reconstruction (arrows (green/red) indicating gamma probe measurements). *Right:* a fast search scan, with much less measurements and covering a much smaller region, used for registration.

3.2 Phantom Experiments

The phantom consists of a plastic cylinder mounted on a rectangular base plate, which also mounts the tracking target with reflective markers. Inside the cylinder three hollow spheres with a radius of 9.86mm each are mounted at mid–height in a triangular fashion with a separation of 4cm each. A Technetium–99m solution is filled into these spheres to simulate radioactive lymph nodes. The three spheres contained an activity of 1.6MBq, 1.6MBq and 1.65MBq. No attenuating medium has been used. One of the spheres could be moved by 2cm away from the other two spheres to simulate deformation during surgery, see Fig. 2 for an illustration.

Three series of experiments were performed on the phantom. *Series 1:* V is a SPECT from the SPECT/CT device, 4 search scans (used as the 1D input b_j for the registration) were performed. *Series 2:* V is an ioSPECT with a search scan performed after reconstruction, repeated 4 times. *Series 3:* V is an ioSPECT scan with a search scan performed after reconstruction, repeated 8 times. All the experiments were performed with the same equipment.

The SPECT V was discretized into $128 \times 128 \times 55$ isotropic voxels of size $4.8mm$. The registration of the SPECT to the tracking coordinate system was performed via the CT that was generated as part of the SPECT/CT acquisition (the SPECT and the CT are automatically co-registered). The ioSPECT used a $50 \times 50 \times 20$ volume for V with isotropic voxels of size $2.5mm$, the scanning protocol is illustrated in Fig. 2. The ioSPECT scans contained an average of 4832 measurements (min. 3576, max. 5445) with a scan duration between 4 and 5 minutes. The search scans all took about 30 seconds and contain 534 measurements on average (min. 513, max. 629). See also Fig. 2 for examples of a full scan versus a search scan. For all the experiments, the actual computation times of the registration process are on the order of seconds.

3.3 Phantom Results

We evaluate several types of errors. First, the accuracy of SPECT and ioSPECT compared to CT for series 1 and 2 is calculated using the euclidean distance

Table 1. Results of the phantom experiments

	accuracy of V compared to CT	ground truth	3D error (average)	2D in–plane error (average)
series 1 (SPECT)	$4.6mm$	CT	$7.5 \pm 1.4mm$	$4.8 \pm 0.6mm$
series 2 (ioSPECT)	$11.9 \pm 6.6mm$	CT	$9.5 \pm 2.3mm$	$7.9 \pm 3.0mm$
series 2 (ioSPECT)	$11.9 \pm 6.6mm$	ioSPECT	$8.1 \pm 6.7mm$	$2.2 \pm 1.2mm$
series 3 (ioSPECT)	n/a	ioSPECT	$7.1 \pm 2.8mm$	$4.2 \pm 2.9mm$

between the centers of mass of the segmented spheres in both modalities. For all series, the 1D–3D registration was performed using the search scan of the moved phantom and compared the result to the ground truth. Here we computed two error measures, the positional euclidean error in 3D, as well as a 2D euclidean error in a plane orthogonal to the surgeon's view. This latter error is more relevant in actual surgeries, as the surgeon is mainly interested in the location of nodes in the plane orthogonal to his viewing direction, since this determines the cutting direction, whereas a node that is located farther away but in the correct cutting direction is less of an issue. All the error types for series 1, 2 and 3 are summarized in Table 1.

For series 3, the search scan data was also tomographically reconstructed (using the same method as for ioSPECT) and then registered to the ioSPECT using a standard SSD–based 3D–3D registration. The 3D average error here was $30.4 \pm 13.8mm$, the 2D average error was $15.2 \pm 7.0mm$. This clearly shows that the search scan data is insufficient for tomographic treatment. In contrast, the error of the proposed 1D–3D registration method was $7.1 \pm 2.8mm$ (3D average error) and $4.2 \pm 2.9mm$ (2D in–plane average error), which is close to the inaccuracies of SPECT and ioSPECT themselves ($4.6mm$ and $11.9 \pm 6.6mm$). It is also interesting to note that the registration error using ioSPECT as ground truth is smaller than the one using CT as ground truth, as both ioSPECT and the search scan for registration use the same physical equipment with the same measurement error sources.

For series 3, we also studied the effect of using less than 500 average search scan data points, i.e. scanning with the probe for even less time. The measurements were thinned out, but still covered the same region, the results are outlined in Table 2. As long as the same area is covered, the registration error does not change significantly, even when just using 10% of the search scan data.

Finally for series 3, a full ioSPECT data scan with an average of 4832 measurements was used as a "search scan", here the average 3D error was $6.3 \pm 2.5mm$, the average 2D error was $3.7 \pm 2.4mm$. This shows that when enough data is available to do a proper tomographic reconstruction, the results will be superior compared to our registration approach. However, in practical situations, having enough data for a reconstruction is not feasible due to the extra duration of the scanning and reconstruction procedure.

Table 2. Results of phantom experiments of series 3 using less search scan data

series 3 (ioSPECT) (ground truth ioSPECT)	no. of measurements	3D error (average)	2D in–plane error (average)
100% search scan	500	$7.1 \pm 2.8mm$	$4.2 \pm 2.9mm$
50% search scan	250	$7.0 \pm 2.7mm$	$4.2 \pm 2.7mm$
33% search scan	165	$6.9 \pm 2.9mm$	$4.2 \pm 3.1mm$
25% search scan	125	$6.7 \pm 2.6mm$	$4.0 \pm 2.7mm$
20% search scan	100	$6.8 \pm 2.5mm$	$4.6 \pm 2.8mm$
10% search scan	50	$6.5 \pm 2.7mm$	$4.5 \pm 2.7mm$

3.4 Experiment with Patient Data

For a more realistic evaluation, we test the registration method on patient data as well. For this an ioSPECT of a breast cancer patient undergoing sentinel lymph node biopsy was used with 7078 measurements reconstructed into a $40 \times 50 \times 25$ volume with isotropic voxels of size $5mm$ (see Fig. 1). The radioactive lymph node was segmented and artificially moved $40mm$ from its original position, similar to the phantom experiment. We perform a search scan of 350 measurements on the patient to register the artificially moved nodes back to their original position. The resulting 3D error with respect to the ioSPECT ground truth is $4.2mm$ and the 2D error is $4.2mm$ as well.

4 Discussion

The accuracy of the proposed registration method is clinically acceptable using ioSPECT as ground truth, since the accuracy of clinical SPECT is between 5mm to 12mm. As one of the aims is to update ioSPECT data, using it as ground truth seems most appropriate. When using pre–operative CT as the ground truth, results suffer, as tracking errors have a higher impact in this scenario. Our evaluation on the patient data set is satisfactory. Even though gathering intra–operative patient data is complex, we plan to investigate further into this direction, in particular analyzing shorter and shorter scan times.

The registration accuracy is limited by the inherent resolution limitations of functional imaging as well as the tracking inaccuracies. However the quality of the input volume used as prior information is also crucial. Intra–operative SPECT scans vary in quality, especially if performed by inexperienced operators. Please note that the flexibility offered by the intra–operative SPECT comes at such a cost. Such compromises within surgical environments are acceptable since the information is provided not to a surgical robotic system, but to an expert surgeon in order to improve his or her performance.

5 Conclusion

To our knowledge this is the first work proposing a 1D–3D registration method for intra–operative functional imaging, which allows registering 3D volumes to

1D measurements with an accuracy suitable for a surgical setting. Our approach provides 3–dimensional image data but requires less readings, and is therefore faster, than a complete scan. Together with upcoming developments of intelligent surgical instruments including integrated detectors this method will enable real–time updates for the guidance of surgeons towards deep–seated targets, resulting in faster and more efficient surgeries and a better quality control.

Acknowledgements. The authors would like to thank Alexander Hartl and Asli Okur for their help in performing the experiments. This work was partially funded by the SFB 824 of the Deutsche Forschungsgesellschaft (DFG), the Graduate School of Information Science in Health (GSISH) and the TUM Graduate School.

References

1. Baumhauer, M., Feuerstein, M., Meinzer, H.P., Rassweiler, J.: Navigation in endoscopic soft tissue surgery: perspectives and limitations. J. Endourol. 22(4), 751–766 (2008)
2. Buck, A.K., Nekolla, S., Ziegler, S., Beer, A., Krause, B.J., Herrmann, K., Scheidhauer, K., Wester, H.J., Rummeny, E.J., Schwaiger, M., Drzezga, A.: SPECT/CT. J. Nucl. Med. 49(8), 1305–1319 (2008)
3. Czernin, J., Benz, M.R., Allen-Auerbach, M.S.: PET/CT imaging: The incremental value of assessing the glucose metabolic phenotype and the structure of cancers in a single examination. Eur. J. Radiol. 73(3), 470–480 (2010)
4. Hartl, A., Ziegler, S.I., Navab, N.: Models of detection physics for nuclear probes in freehand spect reconstruction. In: IEEE Medical Imaging Conference, Knoxville, USA (October 2010)
5. Lyman, G., Giuliano, A., Somerfield, M., Benson III, A.B., Bodurka, D.C., Burstein, H.J., Cochran, A.J., Cody III, H.S., Edge, S.B., Galper, S., Hayman, J.A., Kim, T.Y., Perkins, C.L., Podoloff, D.A., Sivasubramaniam, V.H., Turner, R.R., Wahl, R., Weaver, D.L., Wolff, A.C., Winer, E.P.: American society of clinical oncology guideline recommendations for sentinel lymph node biopsy in early-stage breast cancer. J. Clin. Oncol. 23, 7703–7720 (2005)
6. Valdes Olmos, R.A., Vidal-Sicart, S., Nieweg, O.E.: Technological innovation in the sentinel node procedure: towards 3-d intraoperative imaging. Eur. J. Nucl. Med. Mol. Imaging 37(8), 1449–1451 (2010)
7. Wendler, T., Herrmann, K., Schnelzer, A., Lasser, T., Traub, J., Kutter, O., Ehlerding, A., Scheidhauer, K., Schuster, T., Kiechle, M., Schwaiger, M., Navab, N., Ziegler, S.I., Buck, A.K.: First demonstration of 3-d lymphatic mapping in breast cancer using freehand spect. Eur. J. Nucl. Med. Mol. Imaging (March 2010)

Towards Real-Time 3D US to CT Bone Image Registration Using Phase and Curvature Feature Based GMM Matching*

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

[1] Department of Electrical and Computer Engineering
[2] Department of Orthopaedics
[3] Department of Mechanical Engineering, University of British Columbia, Vancouver, B.C., Canada
{abrou,ilkerh,rafeef}@ece.ubc.ca, pierre.guy@ubc.ca, ahodgson@mech.ubc.ca

Abstract. In order to use pre-operatively acquired computed tomography (CT) scans to guide surgical tool movements in orthopaedic surgery, the CT scan must first be registered to the patient's anatomy. Three-dimensional (3D) ultrasound (US) could potentially be used for this purpose if the registration process could be made sufficiently automatic, fast and accurate, but existing methods have difficulties meeting one or more of these criteria. We propose a near-real-time US-to-CT registration method that matches point clouds extracted from local phase images with points selected in part on the basis of local curvature. The point clouds are represented as Gaussian Mixture Models (GMM) and registration is achieved by minimizing the statistical dissimilarity between the GMMs using an L2 distance metric. We present quantitative and qualitative results on both phantom and clinical pelvis data and show a mean registration time of 2.11 s with a mean accuracy of 0.49 mm.

Keywords: US-CT registration, point cloud matching, local phase features, curvature, Gaussian mixture model registration.

1 Introduction

US has emerged as a desirable intra-operative imaging modality for computer-assisted orthopaedic surgeries, as it is inexpensive, safe and real-time. Despite its many desirable characteristics, it can be difficult for a surgeon to utilize, since B-Mode US images are susceptible to noise, artifacts and a limited field of view. To overcome these deficiencies, US can theoretically be fused with a high resolution image that has a high signal-to-noise ratio, such as pre-operative CT, but in order for US and CT to be successfully used in tandem to visualize bone in orthopaedic surgery, they must be quickly and accurately registered.

Several methods of registering US to CT have been proposed in recent years, but current methods are not sufficiently fast for real-time use. The most widely

* This work was supported in part by NSERC and the Institute for Computing, Information and Cognitive Systems (ICICS) at UBC.

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 235–242, 2011.

used registration method in computer assisted orthopaedic systems (CAOS) to date is the iterative closest point (ICP) algorithm. Methods have been proposed to improve the robustness and speed of the standard ICP method [1,2]. Although relatively successful, ICP exhibits susceptibility to converging to local minima and, therefore, a close initial manual alignment is necessary. Moghari et al. [3] proposed a point-based registration algorithm based on the Unscented Kalman Filter where he achieved more robust and accurate registration results compared to standard ICP. The main drawback of their approach is the need to manually extract bone surfaces from US data. Some researchers have tried to automate the extraction of bone surfaces from US images [4]; however, the techniques were limited to two dimensions (2D). Winter et al. [5] maximized the sum of the overlapping gray values of pre-processed CT bone surfaces and the Three-dimensional (3D) US volume. While this method showed accurate registration results, it assumed a fixed probe orientation for CT pre-processing, which significantly reduces its practicality in real life applications (e.g., in fracture surgery where the US probe needs to be realigned after a fracture reduction). Penney et al. [6] created probability images of the CT and US data and registered these using a normalized cross-correlation metric. The registration results were accurate; however manual segmentation of the CT and US volumes was required to create training sets for the probability images. A more recent method for intensity-based registration involves simulating US images from re-sliced CT data [7] that is updated throughout the registration process, which Gill et al. [8] later extended for registering bone surfaces of the spine. They were able to achieve a mean target 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).

To move closer to the goal of fully automated real-time registration of 3D US to CT volumes, we propose in this paper a registration method based on using local phase information to identify surface points in the US volume, culling the points in both the US and CT volumes based in part on curvature metrics and registering the two point clouds using a GMM technique. We demonstrate improved accuracy and run-time compared to state-of-the-art techniques on pelvic phantom and clinical data.

2 Methods

2.1 Image Pre-processing and Bone Surface Extraction

One of the main challenges of point-based registration of US and CT volumes is extracting a point set from the US data that corresponds with a point set from the CT set. Recently, 3D local phase information has been proposed to extract the surfaces of bones in US volumes [9] using a 3D Log-Gabor filter, $LG(\omega)$ to extract local phase information in the B-Mode US volume:

$$LG(\omega) = e^{-\frac{log(\omega/\omega_o)^2}{2log(\kappa/\omega_o)^2}} \times e^{-\frac{\alpha(\phi_i,\theta_i)^2}{2\sigma_\alpha^2}}. \quad (1)$$

In Equation 1, κ defines the bandwidth of the filter, which determines the frequency specificity. $\alpha(\phi_i, \theta_i)$ is the filter's angle between the azimuth (ϕ) and the elevation (θ) angles. The angular bandwidth is determined by σ. The Log-Gabor filter is built on multiples of the minimum wavelength, λ_{min}, which determine the centre frequency ω_o as $\omega_o = 2/\lambda_{min} \times (\delta)^{m-1}$ for the scaling factor δ and multiple m. The even and odd components, $e_{rm}(x, y, z)$ and $o_{rm}(x, y, z)$, of $LG(\omega)$ are calculated using the real and imaginary responses of the Log-Gabor filter for each point (x, y, z). The 3D phase symmetry (PS) is defined in Equation 2 for each scale (m) and orientation (r):

$$PS_{3D}(x, y, z) = \frac{\sum_r \sum_m \lfloor [|e_{rm}(x, y, z)| - |o_{rm}(x, y, z)|] - T_r \rfloor}{\sum_r \sum_m \sqrt{e_{rm}^2(x, y, z) - o_{rm}^2(x, y, z)} + \epsilon}. \quad (2)$$

Here, T_r is a threshold to account for noise in the US image and ϵ is a small number to avoid division by zero. The bone surface is determined as the maximum PS value along the direction of the US probe.

Compared to the bone surface segmentation of US, CT segmentation is relatively simple. Typically, CT can be segmented using a binary threshold at 200 H.U. Ray-casting is then used to find the bone surface in these binary CT volumes resulting in extracted surfaces that are one voxel thick.

2.2 Point Cloud Selection and Curvature Features

After the bone surface has been extracted from both the US and the CT data, point clouds are created. Our experiments indicate that using between 800 and 1,000 points in each data set produces a reasonable trade-off between a fast run-time and accurate results when matching the volumes using GMMs. Our extraction of both US and CT bone surfaces in Section 2.1 constitutes data reduction from a volume of over 4 million voxels to approximately 10 thousand surface points. To register these volumes in real-time, it is helpful to further significantly reduce the number of points. Therefore, we sub-sample the points extracted in 2.1 to create point clouds for registration, keeping only 5% of the surface points. Though sub-sampling generally preserves the low curvature regions, high curvature regions may be significantly degraded. For example, when matching the iliac spine region of the pelvis using only a sub-sample of the bone surface, registration is accurate and robust with respect to the pelvic table. This is sufficient for 2D registration of the pelvis, but there is little low curvature variability between slices along the iliac spine and this lack of distinguishable features renders the volumes prone to mis-alignment along the crest in 3D registration. To better preserve these salient regions, we use Gaussian curvature, K, to localize high curvature features of the bone surface. K is the product of the principle curvatures κ_1 and κ_2 and is expressed as [10]: $K = \kappa_1\kappa_2 = \frac{eg-f^2}{EG-F^2}$, where $E = ||\mathbf{x}_u||^2$, $F = \mathbf{x}_u \cdot \mathbf{x}_v$ and $G = ||\mathbf{x}_v||^2$ are the coefficients of the first fundamental form, and $e = {}^{(\mathbf{x}_{uu}\mathbf{x}_u\mathbf{x}_v)}/\sqrt{EG-F^2}$, $f = {}^{(\mathbf{x}_{uv}\mathbf{x}_u\mathbf{x}_v)}/\sqrt{EG-F^2}$ and

$g = {}^{(\mathbf{x}_{vv}\mathbf{x}_u\mathbf{x}_v)}/_{\sqrt{EG-F^2}}$ are the coefficients of the second fundamental form for the surface $\mathbf{x}(u, v)$. The top decile of curvature values contribute to the point cloud.

2.3 GMM Registration of US and CT

We registered the US and CT volumes by representing the corresponding point clouds as GMMs and minimizing the distance between the models. GMMs are statistical models representing an entire population using multi-dimensional Gaussian distributions to describe sub-populations, known as components and are represent by the probability density function, $p(\mathbf{x})$. Each component density, ϕ_i, is characterized by its mean, μ_i and its covariance matrix, Σ_i. Using a GMM to represent a point cloud alleviates the need to find point-to-point correspondence between volumes—a major obstacle in many point-based registration methods.

The GMMs are iteratively registered using the L2 similarity metric. The L2 distance has the advantage of having a closed-form solution. The registration algorithm minimizes the L2 cost function of the two GMMs representing the model point cloud, $\mathcal{M}(\mathbf{x}) = p_m(\mathbf{x})$, and the scene, $\mathcal{S}(\mathbf{x}) = p_s(\mathbf{x})$ with a rigid transformation, $\mathbf{T}(\mathcal{M}(\mathbf{x}), \theta, t)$, giving the L2 distance:

$$d_{L2}(\mathcal{M}(\mathbf{x}), \mathcal{S}(\mathbf{x}), \theta, t) = \int (\mathcal{S}(\mathbf{x}) - \mathbf{T}(\mathcal{M}(\mathbf{x}), \theta, t))^2 \mathrm{d}x. \tag{3}$$

It should be noted that the scene model, $\mathcal{S}(\mathbf{x})$, is fixed during the optimization and $\int \mathbf{T}(\mathcal{M}(\mathbf{x}), \theta, t)^2 \mathrm{d}x$ is invariant for rigid transformations, so minimizing the L2 distance in Equation 3 becomes equivalent to solving:

$$\underset{\theta,t}{\operatorname{argmin}} \left[d_{L2}(\mathcal{M}(\mathbf{x}), \mathcal{S}(\mathbf{x}), \theta, t)\right] = \underset{\theta,t}{\operatorname{argmin}} \left[- \int \mathcal{S}(\mathbf{x}) \mathbf{T}(\mathcal{M}(\mathbf{x}), \theta, t) \mathrm{d}x \right]. \tag{4}$$

Given Equation 4 and the formula $\int \phi(\mathbf{x}|\mu_1, \Sigma_1)\phi(\mathbf{x}|\mu_2, \Sigma_2)\mathrm{d}x = \phi(0|\mu_1 - \mu_2, \Sigma_1 + \Sigma_2)$, the closed form expression for the L2 distance between GMMs can be found [11].

2.4 Validation Setup

Phantom Data: In the phantom study we used a Sawbone pelvis #1301 (Pacific Research Laboratories, Inc., Vashon, WA) with 38 1 mm fiducials attached to the surface of the iliac crest and pubic bone. Ten 3D US volumes were acquired of the phantom and were used to create ten US-CT registrations. The US volumes were taken with a G.E. Volusion 730 Expert Ultrasound Machine (GE Healthcare, Waukesha, WI) using a 3D RSP4-12 probe in a water bath. The US phantom volumes were 152x198x148 voxels with an isometric resolution of 0.24 mm. A CT volume was taken with a Toshiba Aquilion 64 (Tustin, CA) and has a resolution of 0.76 mm×0.76 mm×0.3 mm. The pre-processing steps were implemented in

Matlab and the registration was implemented in C++ on a 3.0 GHz Dual Core Intel Xeon CPU using 64-bit Windows 7 and 8 GB of RAM.

Point clouds were created both with and without high curvature features. The fiducial registration error (FRE) was calculated as the average distance between all corresponding fiducial pairs found in both volumes and the surface registration error (SRE) was calculated as the rms distance between the surfaces found in the two imaging modalities.

Each of the ten data sets used for registration as described in Section 2 was comprised of one 3D US volume and a CT volume cropped to the area of interest. The phantom volumes were typically misaligned by as much as 26 degrees and translated by 22mm. For each volume, approximately 800-1000 points were automatically selected on the bone surface. Successful registration should have an error less than 2 - 4 mm for fracture reduction applications in CAOS[12].

Clinical Data: Validation was completed on three US - CT sets acquired from a pelvic fracture patient prior to fracture reduction surgery. The US volumes were acquired using the same setup as described in the phantom study. The CT volume was 512x512x121 voxels and had a resolution of 0.73 mm×0.73 mm×2.0 mm. The SRE and registration run-time were used to quantitatively validate the results.

3 Results and Discussion

Figure 1 shows the results of the pre-processing and registration of the phantom. The resulting surfaces are closely matched, as reflected by the qualitative

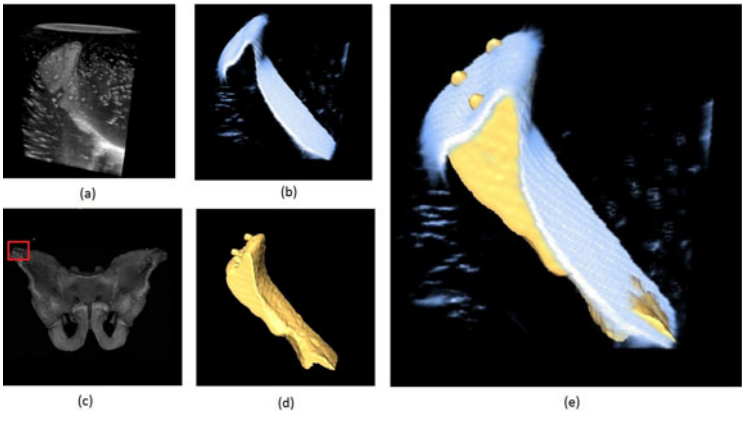


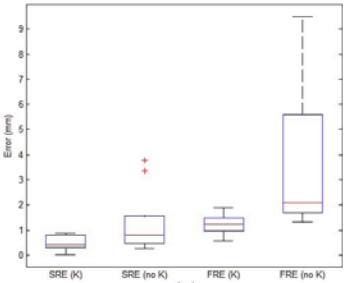
Fig. 1. Qualitative results of the phantom study: (a) is the B-Mode volume of the iliac spine. (b) is the corresponding PS volume. (c) shows the entire CT of the pelvis with the region of interest enclosed in the red box. (d) is the thresholded CT region of interest of the iliac spine. (e) shows the PS volume overlaid on the thresholded CT. Note that the fiducials are visible in both the US and CT volumes.

Phantom	K		no K	
	SRE	FRE	SRE	FRE
Mean (mm)	0.49	1.22	1.30	3.71
Median (mm)	0.42	1.24	0.80	2.10
Std. Dev. (mm)	0.30	0.45	1.26	2.98
Mean Time (s)	2.11		2.12	

(a)

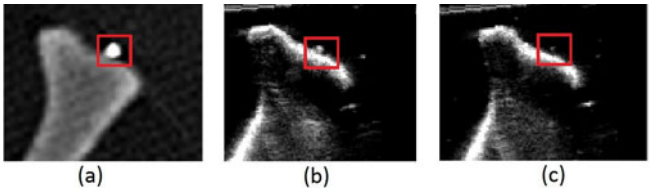
Clinical Scans	K	no$K$
Mean (mm) SRE	0.63	0.84
Mean Time (s)	1.99	1.75

(b)



(c)

Fig. 2. Quantitative results: (a) is a summary of the quantitative phantom results comparing registration completed with and without reinforcing the point clouds with high curvature features, K. (b) is the results of the clinical data study. (c) is the box plot of the phantom results, showing the large standard deviation of the FRE in cases when curvature features are not used.



(a) (b) (c)

Fig. 3. Comparison of registering volumes with and without K: (a) is a slice of a CT volume. (b) is a slice of a US volume that was registered to (a) using K features. (c) is the same slice of an US volume registered without K. Notice the fiducials (red squares) are visible in (a) and (b) and not in (c), due to a misregistration along the iliac crest.

results in Figure 1 (e). The SRE was very low in all tests and averaged 0.49 mm when curvature metrics were used (vs 1.30 mm when not). The FRE noticeably improved when Gaussian curvature features, K, were used to create the point clouds, Figure 2 (c) - 1.22 vs 3.71 mm. The consistent SRE demonstrates the ability of GMM registration to reliably match surfaces; the improved FRE for tests including K indicates that curvature features improve registration when other surface features are too similar. The average run-time of the registration of all tests was less than 2.11 s. It should be noted that the PS surface has been reported to be biased slightly inwards relative to the bone surface (on the order of 0.4 mm) [9], which likely contributes slightly to the FRE.

Qualitative results of the clinical study can be seen in Figure 4. The US volumes were acquired in a region of the iliac spine unaffected by the fracture. The three pairs of registered data sets had a mean SRE of 0.63 mm and run-time of 1.99 s for tests including high curvature points and can be seen in Figure 2.

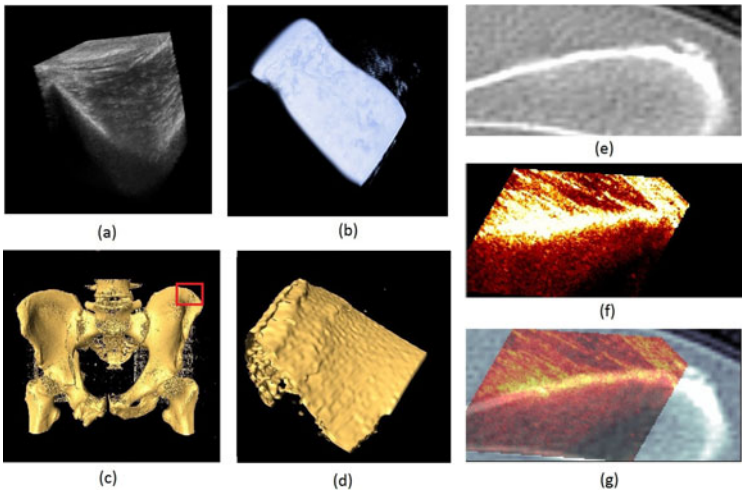


Fig. 4. Qualitative results of the clinical study: (a) is the in vivo B-Mode US and (b) is the PS. (c) is the CT data of a pelvic fracture patient with the region of interest enclosed in the red box. (d) is the thresholded region of interest. (e) and (f) are 2D slices of the CT and US respectively, and (g) is the fused overlaid slice.

4 Conclusion

We proposed a novel method for near real-time 3D US-CT registration using GMM matching of automatically extracted phase and curvature features. We demonstrated the high discriminability of our feature point extraction and have shown that reinforcing highly sub-sampled local phase information with points in high curvature regions improves the robustness and accuracy of the registration without significantly increasing the run-time of the registration. Validating our method on both phantom and real clinical data showed over 80% improvement in run-time compared to state-of-the-art methods [8] with an average accuracy improvement of over 65%. Our method is aimed for prospective US-based OR guidance and thus we will focus our future work efforts on achieving real-time operation by optimizing this algorithm to run on the GPU.

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. Barratt, D.C., Penney, G.P., Chan, C.S.K., Slomczykowski, M., Carter, T.J., Edwards, P.J., Hawkes, D.J.: Self-calibrating ultrasound-to-CT bone registration. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3749, pp. 605–612. Springer, Heidelberg (2005)

3. 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)
4. Kowal, J., Amstutz, C., Langlotz, F., Talib, H., Ballester, M.G.: Automated Bone Contour Detection in Ultrasound B-mode Images for Minimally Invasive Registration in Computer Assisted Surgery an in Vitro Evaluation. The International Journal of Medical Robotics and Computer Assisted Surgery 3(4), 341–348 (2007)
5. Winter, S., Brendel, B., Pechlivanis, I., Schmieder, K., Igel, C.: Registration of CT and Intraoperative 3-D Ultrasound Images of the Spine Using Evolutionary and Gradient-Based Methods. IEEE Transactions on Evolutionary Computation 12(3), 284–296 (2008)
6. 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)
7. 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(5), 577–585 (2008)
8. 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 (August 2010) (in press)
9. 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)
10. Gray, A.: Modern Differential Geometry of Curves and Surfaces with Mathematica, Boca Raton, FL, pp. 373–380 (1997)
11. Jian, B., Vemuri, B.: Robust Point Set Registration Using Gaussian Mixture Models. IEEE Transactions on Pattern Analysis and Machine Intelligence (2011)
12. Phillips, R.: The Accuracy of Surgical Navigation for Orthopaedic Surgery. Current Orthopaedics 21(3), 180–192 (2007)

Robust and Fast Contrast Inflow Detection for 2D X-ray Fluoroscopy

Terrence Chen, Gareth Funka-Lea, and Dorin Comaniciu

Siemens Corporation, Corporate Research,
755 College Road East, Princeton, NJ, USA

Abstract. 2D X-ray fluoroscopy is widely used in computer assisted and image guided interventions because of the real time visual guidance it can provide to the physicians. During cardiac interventions, acquisitions of angiography are often used to assist the physician in visualizing the blood vessel structures, guide wires, or catheters, localizing bifurcations, estimating severity of a lesion, or observing the blood flow. Computational algorithms often need to process differently to frames with or without contrast medium. In order to automate this process and streamline the clinical workflow, a fully automatic contrast inflow detection algorithm is proposed. The robustness of the algorithm is validated by more than 1300 real fluoroscopic scenes. The algorithm is computationally efficient; a sequence with 100 frames can be processed within a second.

Keywords: Contrast detection, fluoroscopy, vessel detection.

1 Introduction

Image-guided interventions have become more and more important for different advanced cardiovascular treatments in recent years. They are widely used for pre-operative planning, intraoperative guidance, and post-operative assessment. In interventional cardiology, 2D X-ray fluoroscopy is still used as the standard imaging modality due to the real time visual guidance it can provide to the physicians. During interventions, it is common to have several angiography acquisitions throughout the entire procedure. An angiography is acquired by injecting a radio-opaque contrast medium into the blood vessel and imaging with X-ray fluoroscopy. Physicians can then visualize the vessels and the blood flow to facilitate the operation.

Although it is rather easy for a physician to tell when the contrast medium appears in a fluoroscopy scene and make judgements accordingly, an automatic contrast inflow detection algorithm is desired for many computer assisted interventions. A few examples include: 1. In a stent enhancement application such as [7], the algorithm needs to discard frames with contrast medium since they could undermine the visibility of stent enhancement. For this purpose, the algorithm needs to determine whether there is contrast injection during the acquisition. If there is, the algorithm has to estimate at which frame the contrast inflow begins to appear and only pre-contrast frames should be processed. 2. In order to

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 243–250, 2011.

generated a DSA-like enhancement on coronary arteries [13], the algorithm has to know when the contrast inflow begins to appear so that masks are collected from frames which are free of contrast medium. 3. In order to improve the overlay of a 3D model when contrast agent is present [6], the system has to know the presence of contrast medium so that it can register the vessel/aorta with a presegmented vessel/aorta model. 4. In order to provide automatic roadmapping [2], the algorithm needs to know when the contrast agent appears in the fluoroscopy to disable the overlay. All these interventional procedures can be streamlined with a robust and automatic contrast inflow detection algorithm such that the physician can obtain desired information with less interactions or constraints.

Robust contrast inflow detection in a large scale is challenging. Little work has been done in the literature. Ideally, sharp changes of the overall intensity histograms along the temporal domain can be used as potential candidates indicating when contrast inflow appears. Nevertheless, this works poorly in practice since moving diaphragm and other darker regions can often appear in cardiac X-ray images and mislead the algorithm. Condurache [2] et al. proposed a method which applies morphological operations and a difference of Gaussians (DoG) filter to generate a vessel map. A threshold is then applied to the histogram of the vessel map to decide whether contrast medium appears.

Based on our experience, the main challenge of automatic contrast inflow detection lies in the high variability of data, especially when the images are often acquired using low dose radiation where the contrast medium is weakly detectable by a traditional filter. Previous methods such as [2] were evaluated only on a limited number of data sets and are with empirically set parameters. They can hardly be generalized to work in different clinical applications.

In this paper, we present a learning-based framework to overcome this problem. Given an input fluoroscopic sequence, the proposed framework first classifies it into a contrast scene (CS), where there is contrast medium injection during the acquisition, or a non-contrast scene (NCS), where contrast medium is not used during the acquisition. For a contrast scene, the algorithm then estimates at which frame the contrast inflow begins to appear in the image. We call this frame the beginning contrast frame (BCF). In order to achieve high robustness against data variability, more than 1300 real fluoroscopic sequences are collected at clinical sites in US, Asia, and Europe. Promising results are obtained.

2 Methodology

The use of the discriminative learning techniques has been proven in many medical imaging applications [1] [11] as well as in interventional applications [13] [10] with high accuracy and efficiency. One of the reasons is that in medical imaging, data variability is often high due to different clinical settings, patients, or dose of radiation. Learning from a large labeled database is one of the most reliable and systematic methods to achieve high robustness against data variability.

Given an input sequence of a fluoroscopic scene, the proposed framework has to first determine whether this is a contrast scene (CS). If it is, the framework

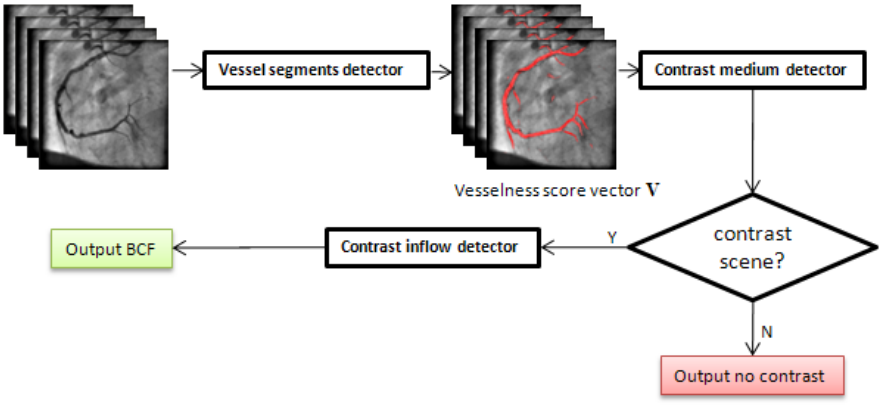


Fig. 1. Workflow of the proposed framework

then estimates the beginning contrast frame. This is more difficult than doing the second step alone, where contrast injection is known for sure to the algorithm.

To solve this problem, we propose a learning-based framework which combines multiple classifiers. Figure 1 illustrates the workflow of the proposed contrast inflow detection framework, where vessel segments detector outputs probability of a patch being inside the vessel or not. The contrast medium detector is a binary classifier to determine whether an input scene is a contrast scene (CS) or a non-contrast scene (NCS). At last, the contrast inflow detector outputs the index of the frame which has the highest probability being the BCF.

2.1 Learning-Based Vessel Detection

T o achieve a robust solution against large data variability, we make use of the discriminative learning technique to facilitate contrast agent detection. Traditional vesselness measurements use different kinds of ridge filters, gradient, edge filters to estimate the vessel-like structures inside an image. Nevertheless, due to the noisy characteristic of the X-ray images, false positives can easily happen. On the other hand, because of preferable low dose radiation in clinics, false negatives on smaller vessels can hardly be avoided. It has been shown that [12] learning based vesselness measurement not only can achieve better performance than conventional filtering based approach [4] but also can be more computationally efficient. For this reason, a learning-based vesselness measurement is applied in our framework. While [12] tried to learn the vesselness in 3D CTA data, we train a learning-based classifier to measure the vesselness for a fluoroscopic image.

To obtain a probability score for a given image to indicate how much contrast (vessel) is present, a binary classifier is implemented as a vessel segment detector, which is learned from a large set of vessel and non-vessel fluoroscopic images. The probabilistic boosting tree (PBT) algorithm [8] is used to learn the classifier. Given an image patch I_{patch}, the PBT classifier calculates the conditional probability that a vessel segment appears in this patch $P(vessel|I_{patch})$. For efficiency,

Fig. 2. Haar wavelet-like feature type examples

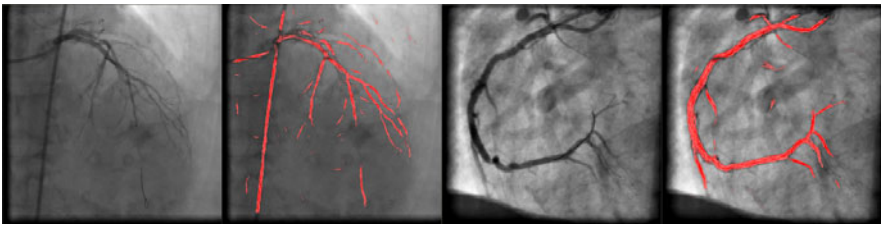


Fig. 3. Input images and the vessel detection results. The darker the red color is, the higher the score is.

the vessel detector is applied to a down-sampled image. Specifically, images are resized to 128×128 for both training and detection. We first apply a set of steerable filters [5] to find ridge points in different orientations in the contrast image. The learned PBT classifier is then applied to each ridge point to compute the probability that the ridge point is from a vessel. A detected vessel region can be defined as the image area where vessel segments are detected with high probabilities. The remaining area of the image contains primarily background structures and is defined as the background region. For vessel segment detection, we choose Haar wavelet-like features generated within a patch size of 8×10. Some examples are shown in Figure 2, which are efficiently calculated using integral image-based techniques [9]. In our implementation, 120000 patches of vessel segments were collected from coronary images to train the PBT classifier. The same number of negative patches were generated automatically by any position which is 10 mm away from the positive patches. The vessel segment detector are tested on 367 images where the coronary vessels were manually annotated. Figure 3 illustrates two examples of the original image and the response of our vessel detector.

2.2 Contrast Medium Detection

With the vessel measure for image, we define a vesselness score vector $\mathbf{V}$ for an input sequence:

$$\mathbf{V} = \{v_1, v_2, \ldots, v_n\}, \tag{1}$$

where n is the number of frames in the sequence. $v_i = \sum P(vessel|I_{patch})$ for all patches in frame i, which represents the vesselness score for the whole image of frame i.

Given the vesselness score vector $\mathbf{V}$, smoothed vectors $\mathbf{G}_{\mathbf{V}_1}, \mathbf{G}_{\mathbf{V}_2}, ..., \mathbf{G}_{\mathbf{V}_m}$ are calculated by applying Gaussian kernels to $\mathbf{V}$ with m different variance σ^2.

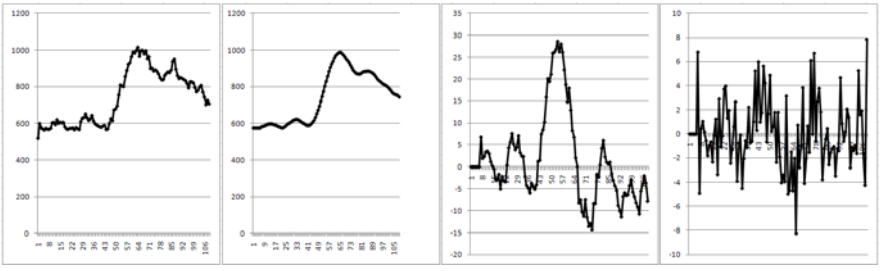


Fig. 4. From left to right: an example of the values of $\mathbf{V}$, and one example of $\mathbf{G_V}$, $\mathbf{F}'$, and $\mathbf{F}''$. The horizontal axis represents the frame index number.

In practice, we set $m = 10$ to include kernel sizes range from 3 to 31 frames. Vectors of the first order and the second order derivatives $\mathbf{F}'_1, \mathbf{F}'_2, ..., \mathbf{F}'_{m+1}$ and $\mathbf{F}''_1, \mathbf{F}''_2, ..., \mathbf{F}''_{m+1}$ for $\mathbf{V}$, $\mathbf{G}_{\mathbf{V}_1}, \mathbf{G}_{\mathbf{V}_2}, ..., \mathbf{G}_{\mathbf{V}_m}$ are then calculated. From each vector of $\mathbf{V}$, $\mathbf{G_V}$s, $\mathbf{F}'$s, and $\mathbf{F}''$s, the following values are calculated as features. The goal is to capture the magnitude and slope of the changes of vesselness measurement throughout the sequence. Let $\mathbf{X}$ be any of the aforementioned vector, $\mathbf{X} \in \{\mathbf{V}, \mathbf{G_V}\text{s}, \mathbf{F}'\text{s}, \mathbf{F}''\text{s}\}$, and let $x_a = min(\mathbf{X}), x_b = max(\mathbf{X}), x_c = median(\mathbf{X}), x_d = mean(\mathbf{X}), x_e = std(\mathbf{X})$. For each $\mathbf{X}$, it is divided into 10 disjoint parts $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_{10}$, where $\sum_i |\mathbf{x}_i| = n$, and $|\mathbf{x}_i| = \frac{n}{10}$. The minimum, maximum, median, mean, and standard deviation are also calculated for all $\mathbf{x}_i$ of each $\mathbf{X}$. At last, absolute differences and ratios between each pair of these values are also added into the feature pool. For example, we find one of the most important features is $|x_b - mean(\mathbf{x}_1)|$ for $\mathbf{V}$, which capture the change of vesselness between the frame with the highest vesselness score and the beginning frames, which are contrast free. Since we only need a binary classifier instead of a probabilistic classifier, a support vector machine (SVM) [3] using radial basis function (RBF) as kernel is then applied to train a binary classifier to decide whether a given sequence is a contrast scene or a non-contrast scene. Figure 4 illustrates an example of $\mathbf{V}$, $\mathbf{G_V}$, $\mathbf{F}'$, and $\mathbf{F}''$.

2.3 Contrast Inflow Detection

If the previous classifier reports that there is no contrast injection in the input scene, the algorithm terminates and outputs NCS to the user. Otherwise, the next task is to find the beginning contrast frame (BCF).

In order to reduce the variation due to different acquisition frame rates, given the vesselness score vector $\mathbf{V}$ of an input sequence, $\mathbf{V}$ is sub-sampled using a standard acquisition frame rate. In our case, 15 fps is chosen as the standard acquisition frame rate. By doing so, we try to normalize the sequence to have a closer blood flow rate and facilitate the detection of BCF. Let $\hat{\mathbf{V}}$ be the sub-sampled vector.

Manual labels of BCF for all the sequences in our database are used as gold standard for a supervised learning. Let x be the labeled position in $\mathbf{V}$ and $\hat{x}$ be

its corresponding position in $\hat{\mathbf{V}}$. $\hat{x}$ is marked as a positive example. For other frame $i \in \{0, ..., |\hat{\mathbf{V}}|\}$, if $i < \hat{x} - 3$ or $i > \hat{x} + 3$, then i is marked as a negative example. 1D Haar features generated from a window size of 11 frames at each example are fed into the PBT classifier [8] to train a classifier $P(f|\hat{\mathbf{V}})$, where f indicates the frame where contrast inflow begins. As a result, the BCF is learned by the classifier:

$$BCF = argmax_{f \in \{0,..,|\hat{\mathbf{V}}|\}} P(f|\hat{\mathbf{V}}). \quad (2)$$

3 Numerical Evaluation

In this section, we present the quantitative evaluation conducted on 1348 fluoroscopy sequences acquired during different interventional procedures, including balloon angioplasty, chronic total occlusion (CTO), intravascular ultrasound catheter pullback, and electrophysiology (EP) procedure. The sequences are acquired in clinical sites in US, Asia, and Europe. Our goal is to collect a database with large data variability.

In the first experiment, we evaluate the performance of the contrast medium detection. In the 1348 sequences, there are 956 contrast scenes and 492 non-contrast scenes. We apply a 4-fold cross validation on our data set, evenly distributed the contrast and non-contrast scenes into 4 parts. For each run, three folds of the data are used for training and the remaining one for testing. The performance is validated by missed detection and false detection. Missed detection denotes that a contrast scene is misclassified as a non-contrast scene and missed detection rate (MDR) is the number of missed detection over the total number of contrast scenes. False detection denotes a non-contrast scene misclassified as a contrast scene. False detection rate (FDR) is the number of false detection over the total number of non-contrast scenes. Table 1 columns A1 and A2 show the average training and testing errors, where the missed detection rate is less than 2% and the false detection rate is less than 5%.

Table 1. Training and testing error of contrast medium detection. The unit of columns B1, B2, B3 is frame number.

	A1. MDR	A2. FDR	B1. Mean	B2. Median	B3. Max
Avg. training error	1.5%	3.3%	0.75	1.82	8
Avg. testing error	1.8%	4.1%	0.9	1.98	11

Although both the missed detection rate and the false detection rate are quite low, we further look into the misclassified scenes and notice the following reasons for misclassification. Most of the missed detection is due to extremely low signal to noise ratio (SNR) or poor contrast visibility and when the scene includes only a major vessel branch with very few branches. In such cases, the vesselness scores of contrast-filled images and non-filled images are hardly distinguishable. Figure 5 (A)-1, and (A)-2 shows one of such scenes. On the other hand, the main

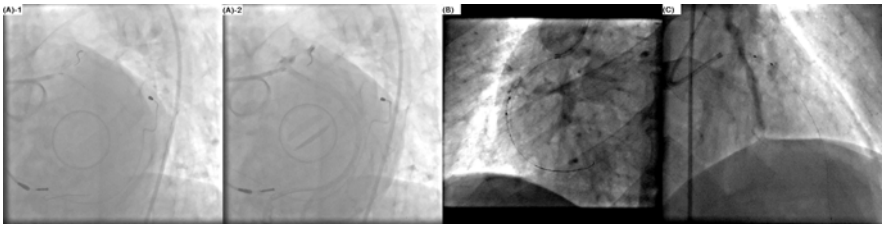


Fig. 5. (A)-1, (A)-2: A missed detection example. (A)-1 shows one frame before contrast injection. (A)-2 shows one of the frames with maximal contrast medium. (B), (C): Two examples of false detection where there is contrast residue moving in the background.

reason for false detection is because in some scenes contrast residue exists in the vessel structures moving in the background, which confuses the detector. Figure 5 (B) and (C) illustrates two examples of false detection. In theory, catheters are hardly distinguishable from the vessels and hence movement of catheters may cause false detection as well. However, this is rarely observed in our experiments.

In the next experiment, we evaluate the accuracy of the beginning contrast frame (BCF), which is measured by the error $|L - D|$, where L is the manual label of BCF, and D is the estimated BCF by the detector. Among the 956 contrast scenes, we again apply a 4 fold cross-validation. Table 1 column B1, B2, and B3 show the average training and testing errors, while Max shows the maximum error across all 4 runs. The mean error is less than 2 frames, which is highly accurate. The accuracy is sufficient for most of real applications. For example, in the stent enhancement application, one can select frames before BCF to generate the enhanced stent.

It is worth noting that another useful output of a contrast inflow detection is a frame with sufficient contrast medium (e.g., presenting one angiography frame to the physician). This can be obtained easily by outputting the frame with the largest v_i in an appropriate $\mathbf{G_V}$ (c.f. $\mathbf{G_V}$ in figure 4)

Lastly, we report the computation time of the proposed framework. Since the vessel detector is only applied to small sub-sampled images (128×128), and the remaining two classifiers are applied to 1D vectors, the computation is very fast. It processes a sequence with 100 frames of original size 1024×1024 within 1 second on a Intel Xeon PC (2.2 GHz) with 3 GB RAM. The average processing time on all the 1348 sequences is about 0.36 seconds, where the average number of frames is 49.3 and the size of images range between 384×384 to 1024×1024.

4 Conclusion

Automatic contrast inflow detection has been found important in many interventional applications. Nevertheless, there is little work in the literature to solve this problem in a large scale. In this paper, we present a learning-based framework which combines three learned classifiers including both SVM and boosting

methods to detect the contrast inflow. Experiments on more than 1300 real fluoroscopic sequences demonstrate that our method is very robust against different clinical settings, dose of radiation, patients, etc. Both missed detection and false detection rates are below 5% and the begin contrast frame can be accurately estimated. In addition, it is with little computational overhead and can be performed within 1 second for most of the fluoroscopic scene. The next step of our work includes but is not limited to extending the current framework for online detection.

References

1. Chen, T., Zhang, W., Good, S., Zhou, S., Comaniciu, D.: Automatic ovarian follicle quantification from 3d ultrasound data using global/local context. In: ICCV (2009)
2. Condurache, A., Aach, T., Eck, K., Bredno, J.: Fast Detection and Processing of Arbitrary Contrast Agent Injections in Coronary Angiography and Fluoroscopy. In: Bildverarbeitung fur die Medizin, Berlin, pp. 5–9 (2004)
3. Cortes, C., Vapnik, V.: Support-Vector Networks. Machine Learning 20 (1995)
4. Frangi, A., Niessen, W., Vincken, K., Viergever, M.: 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)
5. Freeman, W.T., Adelson, E.H.: The Design and Use of Steerable Filters. PAMI 13(9), 891–906 (1991)
6. John, M., Liao, R., Zheng, Y., Nöttling, A., Boese, J., Kirschstein, U., Kempfert, J., Walther, T.: System to Guide Transcatheter Aortic Valve Implantations Based on Interventional C-Arm CT Imaging. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6361, pp. 375–382. Springer, Heidelberg (2010)
7. Koolen, J.J., Vanhet Veer, M., Hanekamp, C.E.E.: A StentBoost image enhancement: first clincal experience. In: MEDICAMUNDI: Clinical Applications (2005)
8. Tu, Z.: Probablistic Bossting-Tree: Learning discriminative models for classfication, recognition, and clustering. In: ICCV (2005)
9. Viola, P., Jones, M.: Robust Real-time Object Detection. IJCV (2002)
10. Wang, P., Chen, T., Zhu, Y., Zhang, W., Zhou, K., Comaniciu, D.: Robust Guidewire Tracking in Fluoroscopy. In: CVPR (2009)
11. Zheng, Y., Barbu, A., Georgescu, B., Scheuering, M., Comaniciu, D.: Four-chamber heart modeling and automatic segmentation for 3d cardiac ct volumes using marginal space learning and steerable features. IEEE Trans. on Med. Img. 27(11), 1668–1681 (2008)
12. Zheng, Y., Loziczonek, M., Georgescu, B., Zhou, S., Vega-Higuera, F., Comaniciu, D.: Machine Learning Based Vesselness Measurement for Coronary Artery Segmentation in Cardiac CT Volumes. In: SPIE Med. Img. (2011)
13. Zhu, Y., Prummer, S., Wang, P., Chen, T., Comaniciu, D., Ostermeier, M.: Dynamic Layer Separation for Coronary DSA and Enhancement in Fluoroscopic Sequences. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5762, pp. 877–884. Springer, Heidelberg (2009)

Generation of 4D Access Corridors from Real-Time Multislice MRI for Guiding Transapical Aortic Valvuloplasties

N.V. Navkar[1,2], E. Yeniaras[2], D.J. Shah[3], N.V. Tsekos[2], and Z. Deng[1]

[1] Computer Graphics and Interactive Media Lab
[2] Medical Robotics Lab, Department of Computer Science, University of Houston, Houston, TX 77004, USA
[3] Methodist DeBakey Heart & Vascular Center, Houston, TX 77030, USA
{nvnavkar,eyeniaras2,ntsekos,zdeng}@cs.uh.edu, djshah@tmhs.org

Abstract. Real-time image-guided cardiac procedures (manual or robot-assisted) are emerging due to potential improvement in patient management and reduction in the overall cost. These minimally invasive procedures require both real-time visualization and guidance for maneuvering an interventional tool safely inside the dynamic environment of a heart. In this work, we propose an approach to generate dynamic 4D access corridors from the apex to the aortic annulus for performing real-time MRI guided transapical valvuloplasties. Ultrafast MR images (collected every 49.3 ms) are processed on-the-fly using projections to extract a conservative dynamic trace in form of a three-dimensional access corridor. Our experimental results show that the reconstructed corridors can be refreshed with a delay of less than 0.5ms to reflect the changes inside the left ventricle caused by breathing motion and the heartbeat.

Keywords: Cardiac Interventions, Real-Time Image-Guided Interventions, 4D Access Corridors, and Magnetic Resonance Imaging.

1 Introduction

The advent of real time image guidance, especially combined with robotic manipulators, may offer new opportunities in interventional medicine. Among the procedures that may benefit from image guidance are intracardiac procedures on the beating heart, such as Transapical Aortic Valve Implantation(TA-AVI) that entail access to the aortic annulus via an apical entrance. Such procedures are usually performed under x-ray fluoroscopy or ultrasound guidance. Three-dimensional ultrasound is commonly used due to its real-time volumetric data collection and lack of ionizing radiation, and it can be combined with robotic systems to synchronize the motion of a device and the heart [7]. Existing literature in the field of image-guided and/or robot-assisted surgeries is vast and inclusive of highly innovative approaches; herein, we only focus on a few most related efforts and it is by no means a comprehensive literature review.

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 251–258, 2011.

In TA-AVI visualization of the Area of Operation (AoO) (i.e., the left ventricle (LV)) is crucial for manual or robot-assisted maneuvering of the interventional tool, that is, constraining interventions inside the dynamic environment of LV without harming healthy tissue. Recently, real-time magnetic resonance imaging (MRI) has emerged as a promising modality for guiding TA-AVI [2,4] since it offers certain advantages [1], such as: (a) a wide range of contrast mechanisms, (b) on-the-fly adjustment of the imaging parameters, (c) an inherent coordinate system, (d) tracking of interventional tools, and (e) lack of ionizing radiation. Robot-assisted, MRI-guided TA-AVI have also been successfully demonstrated [4,9]. For MRI guided intracardiac interventions, several studies have introduced the concept of access corridors, trajectories, and virtual fixtures to reach the targeted anatomy. In addition to imaging speed, another important factor is the extraction and visualization of such access corridors. Ren et al. [8] introduced the concept of dynamic virtual fixtures to assist the operator in minimally invasive procedures in the beating heart. With virtual fixtures, abstract sensory information extracted from preoperative dynamic MRI is overlaid to images to avoid unwanted motion of the interventional tool. Recently, researchers have also reported the generation of dynamic 3D access corridors from preoperative short axis CINE MRI [9] and an efficient algorithm to track the motion of specific anatomical landmarks for guidance [10]. However, both the approaches used CINE MRI that is slow (i.e., requiring several heart beats for one set) and thus inappropriate for real-time guidance. Notably, real-time MRI can reach a speed of 30ms per image [6].

As real-time MRI evolves, new methods can be pursued to improve guidance. In this work, we demonstrates by proper image selection and scheduling, a slow imagining modality (i.e. the MR scanner) can be used in real-time for guidance by generating sufficient information of dynamic interventional environment. We propose a novel algorithm for computing dynamic access corridors from oblique-to-each-other, real-time MRI slices and demonstrate its usefulness for a TA-AVI. The introduced three-step method introduces computational tools for:

- Dynamic MRI by collecting a number of oblique-to-each-other slices with ultrafast MRI prescribed to image particular areas of interest,
- Extraction of the moving LV/endocardium by calculating boundary points from signal intensity projections on those slices, and
- Generation of dynamic access corridor in form of 3D dynamic meshes from those boundary points inside the moving LV.

2 Methodology

Generation of dynamic access corridors is based on two criteria. First, it should ensure that the interventional tool does not harm the moving endocardial wall and other vital structures including mitral valve leaflets, papillary muscles, and chordae tendinae. Second, it should bring the tool to the targeted anatomy (aortic root). In the proposed approach, generation of the dynamic access corridors entails three steps that run in parallel, as shown in Fig. 1, and are described in the following sections.

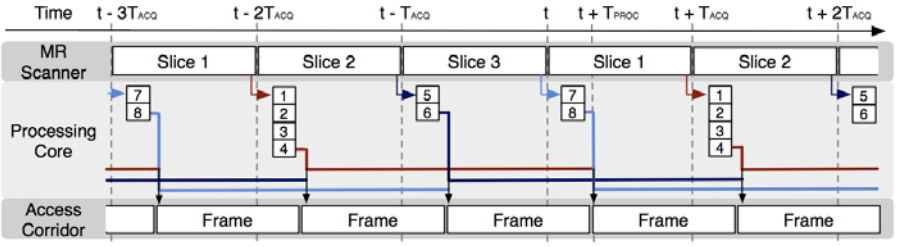


Fig. 1. Illustration of the data collection and processing pipeline: Non-triggered sets of three oblique MR slices are continuously collected (49.3 ms/image) and transferred to the processing core. The core calculates a number of signal intensity projections (4, 2 and 2 for slices 1, 2 and 3 respectively; the numbers 1 to 8 denote the eight signal intensity projections), extracts the boundary points of endocardium and aortic root, organizes and refreshes them (shown by different color lines) to generate a dynamic access corridor. To maintain a uniform frame rate, equivalent to that of MR image collection, data from an individual slice acquisition is retained for the next two acquisition steps until it is refreshed. T_{PROC} is the time lag between MR collection and refreshing the access corridor.

2.1 Collection of Dynamic MR Data

In our particular TA-AVI paradigm, imaging of the AoO is performed by continuously collecting three oblique-to-each-other slices with ultrafast MRI (at $T_{ACQ} =$ 49.3 ms/slice). These slices were selected preoperatively by the interventionalist physician to image particular areas of interest (shown in Fig. 2a). In particular: (1) Slice 1 ($I_1(t)$) is a long axis of view that images the apical region of heart and the aortic and mitral valves, (2) Slice 2 ($I_2(t)$) is also a long axis view that depicts the apical region of the heart, and (3) Slice 3 ($I_3(t)$) is a short axis view prescribed to include the aortic annulus and the LV.

2.2 Extraction of Boundary Points

As each individual MR image is collected, it is sent to a computer via the MR scanner local network, where it is processed to extract the endocardium-LV boundary points $\mathbf{X_{i,j}}(t)$ (i = slice index, j = marker index). The boundary points are calculated from signal intensity projections that correspond to bands PrB_i, with a width W_i and a length L_i (where i = 1 to 8). Those bands are assigned preoperatively by the operator on scout MR runs, to monitor the motion of specific areas of the endocardium. In this case, the operator selects four projection bands on slice $I_1(t)$ and two each on slices $I_2(t)$ and $I_3(t)$ (shown in Fig. 2a). The selection of the projection bands is arranged so that PrB_1, PrB_2, PrB_5, and PrB_6 depict the motion of the endocardium, whereas PrB_3, PrB_4, PrB_7, and PrB_8 capture the motion of aortic root.

To calculate the boundary points in the coordinate system of the MR scanner, the following algorithm is used. For each projection band, a local 2D coordinate system is defined on the slice with its origin set at the center of the projection

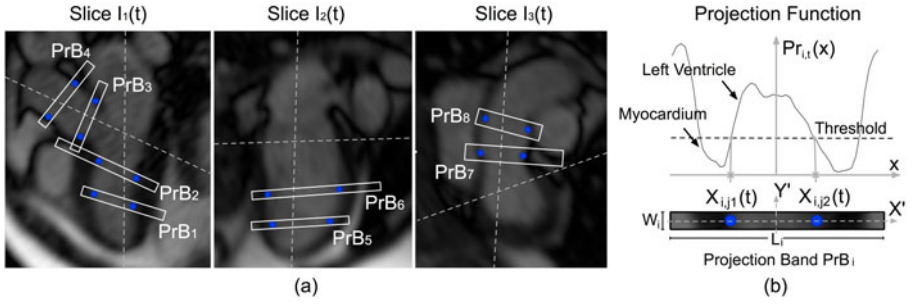


Fig. 2. (a) Example of the three slices ($I_1(t)$, $I_2(t)$ and $I_3(t)$) and the eight operator prescribed projection-bands (PrB_1 to PrB_8). The dashed lines on any one of the slices show the intersection of other two slices, and the blue dots correspond to the boundary points. (b) Representative Projection $Pr_{i,t}(x)$ generated from the projection-band.

band PrB_i and the axes parallel and orthogonal to its length (Fig. 2b). A projection profile $Pr_{i,t}(x)$ is generated along the local X-axis of the band. For each value of x, $Pr_{i,t}(x)$ represents the averaged signal intensity along its width, i.e. from $(x, W_i/2)$ to $(x, -W_i/2)$ at time t.

In the TrueFISP images, the myocardium exhibits lower signal intensity as compared to the LV, and on the projection signal intensity appears as a deep. To identify this deep, a threshold is applied and an algorithm traverses the projection function along the X-axis starting from its center and moving along both directions. The threshold is also selected manually from the scout MR scans using a custom GUI gadget. A point is marked on the X-axis if the value of $Pr_{i,t}(x)$ falls below the threshold. After this step, two boundary points $\mathbf{X_{i,j1}}(t)$ and $\mathbf{X_{i,j2}}(t)$ are extracted on the slice i at time t (indexing of the boundary points with respective projection bands is given in Table 1). It takes 9 to 12 minutes for cardiac MR technicians to set up the initial parameters for the pipeline.

2.3 Generation of 4D Dynamic Access Corridor

In this step, the dynamic access corridor at time t is represented with a triangular mesh $\mathbf{M(t)}$ and is generated from the boundary points using a two-stage mesh reconstruction process (Fig. 3a). In the first stage, a coarse structure is created using the boundary points. The eight boundary points that correspond to the motion of the endocardium, extracted from $\{PrB_i\}$ (i = 1, 2, 5 and 6), are interconnected to define a coarse mesh $\tilde{\mathbf{M}}_{\mathbf{E}}(t)$. Similarly, a coarse mesh $\tilde{\mathbf{M}}_{\mathbf{A}}(t)$ is defined for the aortic root from the remaining boundary points, extracted from $\{PrB_i\}$ (i = 3, 4, 7 and 8). The region between the two meshes is created by using Kochanek-Bartels curves [3]. As an example (shown in Fig. 3a) , curve $\mathbf{c_{1,t}}(u)$ is defined by the boundary points $\mathbf{X_{1,1}}(t)$, $\mathbf{X_{1,3}}(t)$, $\mathbf{X_{1,5}}(t)$, $\mathbf{X_{1,7}}(t)$ and is interpolated between the points $\mathbf{X_{1,3}}(t)$ and $\mathbf{X_{1,5}}(t)$. Curves $\mathbf{c_{2,t}}(u)$, $\mathbf{c_{3,t}}(u)$, and $\mathbf{c_{4,t}}(u)$ are defined in a similar manner. It is noteworthy that the tangential properties of the curve can be altered to adjust the deflection of the access corridor from the endocardium to the aortic root.

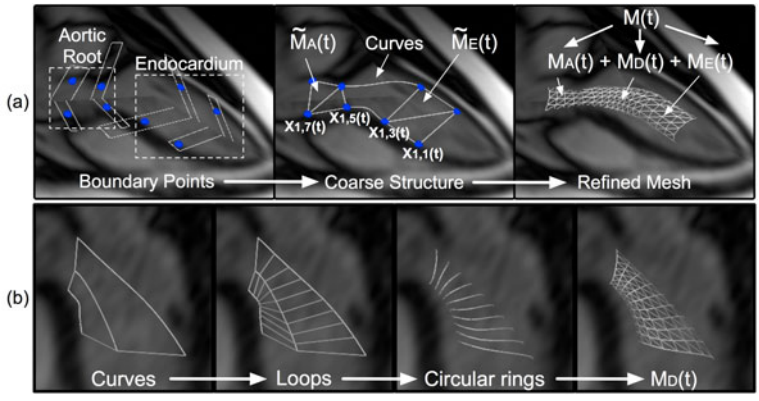


Fig. 3. (a) Generation of dynamic access corridor $\mathbf{M(t)}$ from boundary points (overlaid on slices $I_1(t)$, $I_2(t)$ and $I_3(t)$). (b) Generation of deflection mesh $\mathbf{M_D}(t)$ from the curves (overlaid on slice $I_1(t)$).

The second stage refines the above coarse mesh structure. Points at regular intervals on the curves $\mathbf{c_{1,t}}(u)$, $\mathbf{c_{2,t}}(u)$, $\mathbf{c_{3,t}}(u)$, and $\mathbf{c_{4,t}}(u)$ are interconnected to form intermediate loops, which are further subdivided to get circular rings [5]. These circular ring are then interconnected to form a deflection mesh $\mathbf{M_D}(t)$ (Fig. 3b). The coarse meshes $\tilde{\mathbf{M}}_\mathbf{A}(t)$ and $\tilde{\mathbf{M}}_\mathbf{E}(t)$ are also subdivided to get finer meshes $\mathbf{M_A}(t)$ and $\mathbf{M_E}(t)$, respectively [5]. Since for the three meshes $\mathbf{M_D}(t)$, $\mathbf{M_A}(t)$, and $\mathbf{M_E}(t)$ we use the same subdivision scheme (with the same number (n=2) of iterations), the three meshes can be stitched together to form the final mesh $\mathbf{M(t)}$ without altering the positions of the boundary vertices.

Our refinement process is conservative in nature as the mesh $\mathbf{M(t)}$ is confined within the region defined by the coarse meshes $\tilde{\mathbf{M}}_\mathbf{A}(t)$ and $\tilde{\mathbf{M}}_\mathbf{E}(t)$ and the curves $\mathbf{c_{i,t}}(u)$ (where i = 1 to 4). The region defined by the coarse meshes and the curves is confined by the anatomical structures. For every time frame, the final mesh $\mathbf{M(t)}$ will always have the same number of vertices and faces. Thus, the vertices on the mesh could be indexed based on the nearby anatomical structure. This allows the corridor to provide an adaptive guidance mechanism for an interventional region. The total time required for computing the mesh from the boundary points is given by T_{MESH}.

3 Experimental Studies and Discussion

The proposed approach was experimentally tested and validated on a Siemens 1.5T Avanto MR scanner. Multislice non-triggered and free-breathing imaging was performed with a true fast imaging with steady-state precession (TrueFISP), at a T_{ACQ} = 49.3 ms per slice (Pixel Spacing: 1.25x1.25; FOV: 275x400; TR: 49.3ms; TE: 1ms; Matrix: 320x200; and Slice Thickness: 6mm). The computational core was implemented on a dedicated PC (Intel 3.2GHz processor with

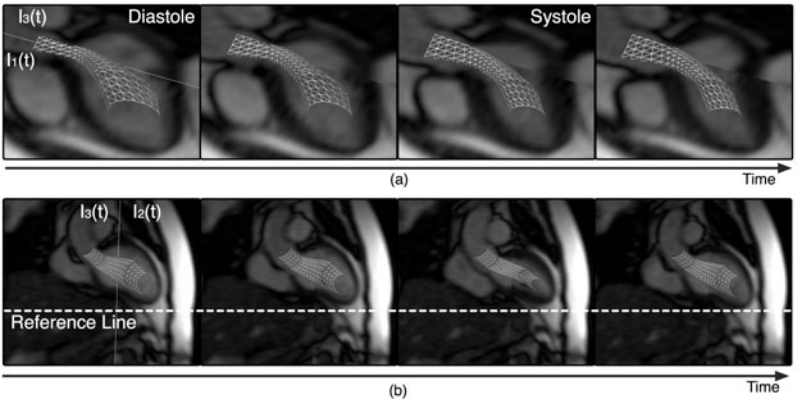


Fig. 4. Four selected frames from the stream generated by the computational core showing the dynamic access corridor overlaid to the real time MR images (a) $I_1(t)$, $I_3(t)$ and (b) $I_2(t)$ and $I_3(t)$. In (b) the dashed line is included for appreciation of motion caused by breathing.

9GB RAM). Fig. 4 shows an example outcome of the access corridor superimposed onto two slices (out of three) for the sake of clarity. The deformation of the access corridor secondary to heart beating, as well as its relative motion due to free breathing, can be appreciated in those frames. For all the time frames in our studies, we found the access corridor never collided with the endocardium and aortic root depicted on the real-time MR images. The corridor acts as a 'base mesh' and could be further processed as per the needs of intervention.

With the interleaved multislice MRI, each individual slice is refreshed every $3T_{ACQ}$. Thus, one interesting question is to what degree this reduced (one third) refreshing rate may result in loss of information. To assess this, we performed experiments that continuously collected only one slice (every 49.3 ms) and extracted the boundary points. Variations with time in the distance of the boundary points from the origin $\|\mathbf{X_{i,j}}(t)\|$ (measured in the local coordinate system of the projection band) were calculated. From those data, two signals were generated (shown in Fig. 5a): the original (i.e. complete series) signal and one that sampled the original every $3T_{ACQ}$. The difference between the two signals was used to calculate the error due to interleaving (shown in Fig. 5b). The mean error E_M and the correlation coefficient κ of the two signals for a period of 9 seconds is shown in Table 1. For all the boundary points, E_M stays below two pixel spacing.

Table 1 summarizes our analysis results reporting the time required to compute each projection band ($T_{Pr,i}$) and the time lag (T_{PROC}) between MR collection and refreshing the access corridor. Specifically, T_{PROC} is computed as max $(\sum_{i=1}^{4} T_{Pr,i}, \sum_{i=5}^{6} T_{Pr,i}, \sum_{i=7}^{8} T_{Pr,i}) + T_{MESH}$. In the experiment (parameters described in Table 1), the measured T_{PROC} was 0.261ms (i.e., 1/184 of T_{ACQ}); and 0.194 ms (T_{MESH}) was used for the meshing step. In general, for different configurations T_{PROC} stayed below 0.5 ms.

Table 1.

Slice	Projections	L_i(mm)	W_i(mm)	$T_{Pr,i}$(ms)	Points	E_M(mm)	κ
$I_1(t)$	PrB_1	55.13	6.13	0.018	$\mathbf{X_{1,1}}(t)$	1.22	0.83
					$\mathbf{X_{1,2}}(t)$	1.07	0.84
	PrB_2	70.40	6.09	0.022	$\mathbf{X_{1,3}}(t)$	1.03	0.81
					$\mathbf{X_{1,4}}(t)$	1.57	0.87
	PrB_3	47.05	5.93	0.013	$\mathbf{X_{1,5}}(t)$	1.41	0.78
					$\mathbf{X_{1,6}}(t)$	1.89	0.64
	PrB_4	49.69	6.02	0.014	$\mathbf{X_{1,7}}(t)$	1.25	0.81
					$\mathbf{X_{1,8}}(t)$	1.12	0.83
$I_2(t)$	PrB_5	62.84	5.07	0.021	$\mathbf{X_{2,1}}(t)$	1.36	0.83
					$\mathbf{X_{2,2}}(t)$	1.18	0.84
	PrB_6	82.53	5.07	0.027	$\mathbf{X_{2,3}}(t)$	1.79	0.83
					$\mathbf{X_{2,4}}(t)$	1.90	0.74
$I_3(t)$	PrB_7	64.96	18.94	0.043	$\mathbf{X_{3,1}}(t)$	0.60	0.92
					$\mathbf{X_{3,2}}(t)$	0.43	0.95
	PrB_8	43.45	10.23	0.017	$\mathbf{X_{3,3}}(t)$	0.71	0.90
					$\mathbf{X_{3,4}}(t)$	0.80	0.84

Future work can be geared towards optimizing MR collection in two directions. First, while in this work the flow of information is one-way (i.e., from the MR scanner to the processing core), real-time feedback to the MR scanner can be added to adjust the orientation of the imaging planes on-the-fly based on projections. Second, 2D imaging can be substituted with the collection of actual MR projections, as is the case with navigator echoes (with a 90-180 MR pulse sequence to select a column through the tissue and sample the actual projection from the MR signal). This can further speed up data acquisition (since a complete image is not collected) and, in addition, improve the SNR (since a read out projection is collected instead of an image).

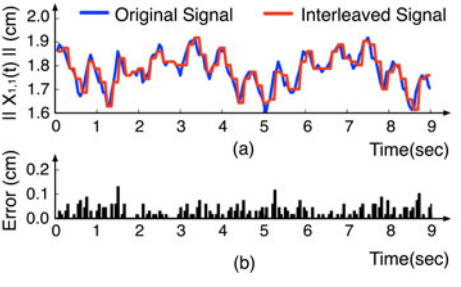


Fig. 5. (a) Comparison of the time-delay effect. The single slice signal is sampled at 20.28Hz (i.e. 49.3 ms/slice) and the multislice at 6.76Hz (i.e. 147.9 ms/slice). (b) The error between single and multislice (n=3) acquisition. The "Error" (Y axis) is the absolute value of the difference between the two signals.

4 Conclusion

This work proposes an approach to generate 4D access corridors for performing interventions in the beating heart, from a non-triggered continuously acquired set of oblique-to-each-other MR images. This approach is largely motivated by the challenge associated with the inherent low sensitivity of MRI modality that prevents collecting high SNR, and often high CNR, images in real-time. The reconstructed corridor is virtually refreshed with the same speed as the individual

MR slices are collected (in this case, 49.3 ms/image) with a delay of less than 0.50 ms. This dynamic corridor can be used, without or with an additional safety margin (for a more conservative approach), for visual servoing, image-based robot control, or force-feedback-assisted manual control.

Acknowledgments. This work was supported by the National Science Foundation (NSF) award CPS-0932272. All opinions, findings, conclusions or recommendations expressed in this work are those of the authors and do not necessarily reflect the views of our sponsors.

References

1. Guttman, M.A., Ozturk, C., Raval, A.N., Raman, V.K., Dick, A.J., DeSilva, R., Karmarkar, P., Lederman, R.J., McVeigh, E.R.: Interventional cardiovascular procedures guided by real-time mr imaging: an interactive interface using multiple slices, adaptive projection modes and live 3d renderings. Journal of Magnetic Resonance Imaging 26(6), 1429–1435 (2007)
2. Horvath, K.A., Mazilu, D., Kocaturk, O., Li, M.: Transapical aortic valve replacement under real-time magnetic resonance imaging guidance: experimental results with balloon-expandable and self-expanding stents. European Journal of Cardiothoracic Surgery 39(6), 822–828 (2010)
3. Kochanek, D., Bartels, R.H.: Interpolating splines with local tension, continuity, and bias control. SIGGRAPH Computer Graphics 18(3), 33–41 (1984)
4. Li, M., Kapoor, A., Mazilu, D., Horvath, K.A.: Pneumatic actuated robotic assistant system for aortic valve replacement under mri guidance. IEEE Transaction on Biomedical Engineering 58(2), 443–451 (2011)
5. Loop, C.: Smooth subdivision surfaces based on triangles (master thesis) department of mathematics, university of utah (1987)
6. McVeigh, E.R., Guttman, M.A., Kellman, P., Raval, A.N., Lederman, R.J.: Real-time, interactive mri for cardiovascular interventions. Academic Radiology 12(9), 1121–1127 (2005)
7. Mebarki, R., Krupa, A., Collewet, C.: Automatic guidance of an ultrasound probe by visual servoing based on B-mode image moments. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 339–346. Springer, Heidelberg (2008)
8. Ren, J., Patel, R.V., McIsaac, K.A., Guiraudon, G., Peters, T.M.: Dynamic 3-d virtual fixtures for minimally invasive beating heart procedures. IEEE Transaction on Medical Imaging 27(8), 1061–1070 (2008)
9. Yeniaras, E., Deng, Z., Syed, M.A., Davies, M., Tsekos, N.V.: Virtual reality system for preoperative planning and simulation of image guided intracardiac surgeries with robotic manipulators. In: Studies in Health Technology and Informatics, vol. 163, pp. 716–722. IOS Press, Amsterdam (2011)
10. Zhou, Y., Yeniaras, E., Tsiamyrtzis, P., Tsekos, N., Pavlidis, I.: Collaborative tracking for MRI-guided robotic intervention on the beating heart. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6363, pp. 351–358. Springer, Heidelberg (2010)

Automatic Trajectory Planning of DBS Neurosurgery from Multi-modal MRI Datasets

Silvain Bériault, Fahd Al Subaie, Kelvin Mok, Abbas F. Sadikot, and G. Bruce Pike

McConnell Brain Imaging Centre, Montreal Neurological Institute,
3801 University Street, Montreal, Quebec, H3A 2B4, Canada
silvain.beriault@mail.mcgill.ca

Abstract. We propose an automated method for preoperative trajectory planning of deep brain stimulation image-guided neurosurgery. Our framework integrates multi-modal MRI analysis (T1w, SWI, TOF-MRA) to determine an optimal trajectory to DBS targets (subthalamic nuclei and globus pallidus interna) while avoiding critical brain structures for prevention of hemorrhages, loss of function and other complications. Results show that our method is well suited to aggregate many surgical constraints and allows the analysis of thousands of trajectories in less than 1/10th of the time for manual planning. Finally, a qualitative evaluation of computed trajectories resulted in the identification of potential new constraints, which are not addressed in the current literature, to better mimic the decision-making of the neurosurgeon during DBS planning.

Keywords: Deep brain stimulation, Parkinson's disease, image-guided neurosurgery, automatic planning.

1 Introduction

Over the past decade, Deep Brain Stimulation (DBS) has become a valuable surgical treatment to severe Parkinson's Disease (PD) – a neurodegenerative disease that affects 1% of population over 60 years of age [1] and over one million people in North America [2]. DBS consists of the surgical insertion of stimulation electrodes, in specific nuclei of the basal ganglia circuitry, programmed to reduce PD symptoms. These electrodes are inserted via minimally invasive neurosurgery using precise image guidance from a neuro-navigation platform.

Before the operation, the neurosurgeon undertakes a preoperative planning procedure. The goal is to examine the patient's imaging data to determine: i) precise target locations where to implant the DBS electrodes and ii) safe linear trajectories, from the surface of the head to the targets, that avoid critical structures of the brain to prevent hemorrhages, loss of function and other complications.

Trajectory planning is normally done by manual inspection of a single anatomical MRI dataset using visualization tools offered by commercial neuronavigation platforms. The surgeon empirically searches for a safe trajectory that avoids several critical structures such as: i) the ventricles, ii) deep sulci, iii) large blood vessels, and iv) critical motor and sensory cortex. However, only few trajectories can be thoroughly analyzed in a reasonable amount of time therefore yielding subjective and sub-optimal

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 259–266, 2011.

planning. There has been recent interest in the design of automatic planning algorithms to allow the rapid analysis of many more trajectories across multi-modal imaging datasets. With this new paradigm, mimicking the decision-making process of neurosurgeons poses an important challenge.

Proposed methods for automatic planning can be distinguished based on: i) the level of automation, ii) the choice of critical structures and their representation, iii) the trajectory analysis criteria and final scoring, and iv) the overall ease-of-use. Some software tools [3-4] were proposed to simulate the surgical insertion and automatically detect intersected structures, but the entry point selection remains empirical. Many methods encode critical structures once in an atlas [5-7] although recent methods favor direct segmentation in the native patient datasets to better account for inter-subject variability, especially for important structures such as sulci and blood vessels. Very relevant to our work, the method of Brunenberg *et al.* [8] returns many valid trajectories with no further ranking. In the work of Shamir *et al.* [9], trajectories are ranked separately according to either a maximal risk or a sum or risks criteria with no further aggregation. The method of Essert *et al.* [10] automatically ranks trajectories according to several weighted surgical constraints defined by the neurosurgeon using an elaborated extensible markup language (xml) schema. They compare automatic and manual planning according to the final aggregated score, rather than showing how each individual constraint, taken separately, is optimized.

In this work, we present an automatic path planning framework that incorporates several key improvements at every stage of the process: from MRI acquisition to automatic trajectory selection. First, our approach takes advantage of the most recent advances regarding venous and arterial blood vessel imaging with the use of susceptibility weighted imaging (SWI) [11-12] and time-of-flight (TOF) [13] protocols. Second, our trajectory analysis software can handle binary and fuzzy segmentation datasets without the use of a global threshold or of lengthy, iterative, post-processing. Third, our framework meaningfully aggregates several, easy-to-configure, clinical criteria into a single trajectory ranking.

2 Method Overview

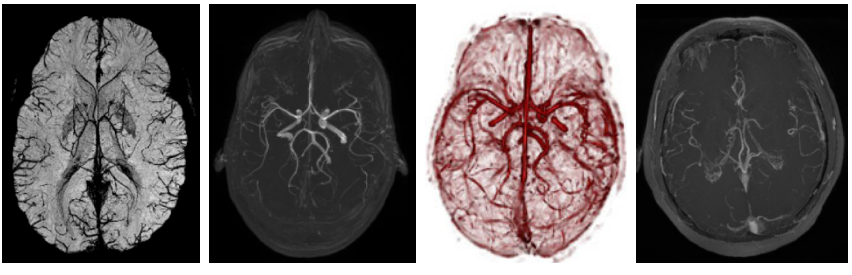
Our framework performs multi-modal analysis of patient data to determine the most suitable trajectory to a DBS target according to a set of constraints defined by the surgeon. This section provides detailed information about the multi-modal MRI acquisition protocol, the definition of relevant surgical constraints, and the implementation of the trajectory planning algorithm.

2.1 MRI Acquisition

A multi-modal MRI acquisition protocol is performed on a 3T Siemens TIM Trio with a 32-channel coil. First, a sagittal T1w anatomical scan of the entire head with 1x1x1-mm resolution is obtained using a 3D magnetization-prepared rapid gradient-echo (MP RAGE) sequence (TR=2300ms, TI=900ms, TE=2.98ms, α=9°). Second, a transverse SWI dataset of the brain with 0.5x0.5x1-mm resolution is obtained using a fully flow compensated 3D gradient echo sequence (TR=34ms, TE=20ms, α=12°,

BW=120Hz/px). Third, an MRA dataset is obtained with 1x1x1-mm resolution using a 3D multi-slab TOF (4 slabs, 44 slices/slab, transverse acquisition, TR=22ms, TE=3.85ms, α=18°). These three datasets are acquired in a single session and are aligned by linear registration (rigid body, 6 parameters). Furthermore, the T1w dataset is aligned to the ICBM-152 atlas by non-linear registration [14].

The SWI and TOF sequences provide dense visualization of venous and arterial vasculature (see Fig. 1a-c) without requiring the injection of gadolinium, which adds significant cost to each study and has some associated medical risks. In particular, the SWI sequence (Fig.1a) can image very small veins in comparison to a typical gadolinium protocol (e.g. Fig. 1d). While this dense SWI-TOF protocol allows for a safer trajectory to be planned, the high density makes manual planning a very laborious and challenging task.



(a) (b) (c) (d)

Fig. 1. SWI-TOF protocol for dense visualization of venous and arterial blood. (a) 30-mm minimum intensity projection (mIP) of an SWI dataset. (b) Maximum intensity projection (MIP) of a TOF dataset. (c) 3D rendering of combined SWI-TOF dataset. (d) 30-mm MIP of a PD patient's clinical scan (T1w with gadolinium).

2.2 Surgical Constraints Definition

Our automatic trajectory planning framework is governed by a set of customizable surgical constraints obtained from multiple interview sessions with a senior neurosurgery resident and an experienced neurosurgeon. These constraints are summarized below. As discussed in Section 4, this list can be further extended and customized as new constraints are found to interplay with the surgeon's decision-making.

1. *Avoid critical cortex area.* To reduce the risk of introducing new neurological deficits, the surgeon typically selects an entry point within the superior frontal lobe, anterior to the primary motor cortex and posterior to the hairline. To mimic this behavior, entry points are limited to a surgeon-chosen region-of-interest (ROI) defined once on the ICBM-152 atlas.
2. *Avoid crossing the midline.* The midline is avoided because it is filled with CSF, it encompasses the thick mid-sagittal sinus vessel and it implicitly eliminates long and risky paths. In our implementation, any entry points on the opposite side of the target are immediately discarded.

3. *Avoid ventricles.* Ventricles are avoided to prevent CSF leaks and to remain in control of the lead. Ventricles are automatically segmented from the T1w dataset using a standard tissue classification method [14].
4. *Avoid blood vessels.* Vessels are avoided to prevent intra-operative complications such as hemorrhages. Veins (SWI dataset) and arteries (TOF dataset) are segmented using a fuzzy vesselness measure [15].
5. *Avoid sulci.* Sulci are avoided because they contain many small vessels that may not be well depicted by current MRI protocols. Sulci are segmented from the T1w dataset using a standard tissue classification method [14].

2.3 Automatic Trajectory Planning

The trajectory planning algorithm consists of analyzing every trajectory linking a set of entry points (on the head surface) to the target. Similarly to work of Essert *et al.* [10], our analysis is conducted in two passes. A first pass quickly eliminates any trajectory that crosses a critical structure at an unsafe distance. With the remaining trajectories, a second pass optimizes the distance to all critical structures simultaneously. Pseudo-code for the second pass is shown at the end of this section.

In other related work [6, 8-10], trajectory optimization is based on a pre-calculated distance map that encodes the minimal distance of a voxel from a critical structure. This method is computationally efficient because the distance map is computed only once and applied to every trajectory. However, this technique only consider the distance to the closest critical structure, therefore giving too much importance to false positive voxels over large clusters representing true critical structures (see Fig. 2 for an example). Furthermore, a distance map can be calculated only on binary segmented datasets. Instead, we represent each trajectory as a cylinder of interest with a N-mm radius (N = 5-mm) i) to account for the dimension of the insertion tool and the precision of patient-to-image registration, and ii) to exclude critical structures already at a safe distance (above 5-mm) without any further processing. With this technique, a distinction can be made between the thick structure of Fig. 2a and the isolated structure of Fig. 2b because, locally, more foreground voxels would intersect the cylinder of interest.

A risk value is given to every voxel inside the cylinder according to the distance from the cylinder's centerline and the voxel value (for fuzzy datasets). For example, the associated risk of a fuzzy vesselness voxel is determined by a combination of: i) its distance from trajectory`s centerline and ii) its vessel-likeliness value. This allows the extraction of many statistical parameters (e.g. $risk_{max}$, $risk_{sum}$, $risk_{mean}$, $risk_{median}$, etc). In this prototype, the maximal risk ($risk_{max}$) and sum of risks ($risk_{sum}$) are extracted and normalized, using a histogram analysis, to a [1 100] range for comparison with other surgical constraints [10]. A final trajectory score is computed by aggregating the $risk_{max}$ and $risk_{sum}$ parameters for all surgical constraints using a weighted cost function (see Algorithm 1). The weights are chosen by the neurosurgeons to represent the relative importance of each constraint. A greater weight is usually given to the $risk_{max}$ criterion because almost hitting a critical structure once is more severe than approaching the same critical structure multiple times at a safer distance [8]. On the other hand, the sum criterion is useful because it distinguishes among cases where a trajectory approaches a critical structure once or multiple times along the path [9].

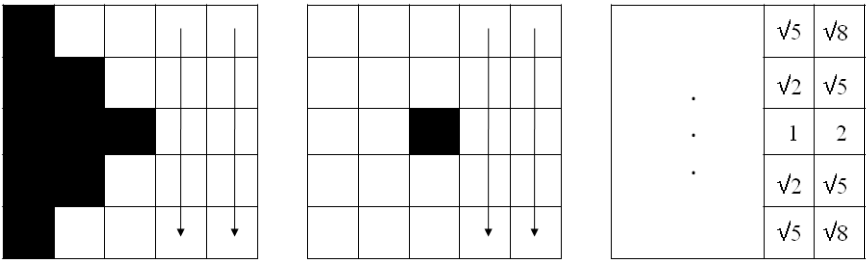


Fig. 2. Distance map calculation fails to distinguish between (a) a true critical structure and (b) an isolated false positive voxel as they both result in the same distance map (c) for the two rightmost columns

Algorithm 1. Trajectory optimization (second pass)

```
Input: ePnts[p];     % list of entry points
Input: tPnt;         % the target point
Input: cs[i];        % list of critical structures
Input: wmax[i];      % weights for each cs (riskMax)
Input: wsum[i];      % weights for each cs (riskSum)
Output: fScores[p]; % final scores for each trajectory
for each p in ePnts
  cylinder = CalcCylinderOfInterest(p, tPnt);
  for each i in cs
    for each j in cylinder
      costs[p][i][j] = CalcCost(cylinder[j], cs[i]);
    riskMax[p][i] = FindMax(costs[p][i][:]);
    riskSum[p][i] = FindSum(costs[p][i][:]);
for each i in cs
  NormalizeToRange(riskMax[:][i], [1 100]);
  NormalizeToRange(riskSum[:][i], [1 100]);
for each p in ePnts
  fScores[p] = 0;
  for each i in cs
    fScores[p]+=riskMax[p][i]*wmax[i];
    fScores[p]+=riskSum[p][i]*wsum[i];
Sort(fScores)
```

3 Experimental Results

Our automatic trajectory planning tool was evaluated by a senior neurosurgery resident and an experienced neurosurgeon. Multi-modal MRI datasets were acquired on two healthy subjects and one patient (M/47 yrs) with severe PD rigidity symptoms. For all subjects, four DBS targets were manually identified by the surgeons – within the two subthalamic nuclei (STN) and the two globus pallidus interna (GPi) – for a total of 12 planning experiments.

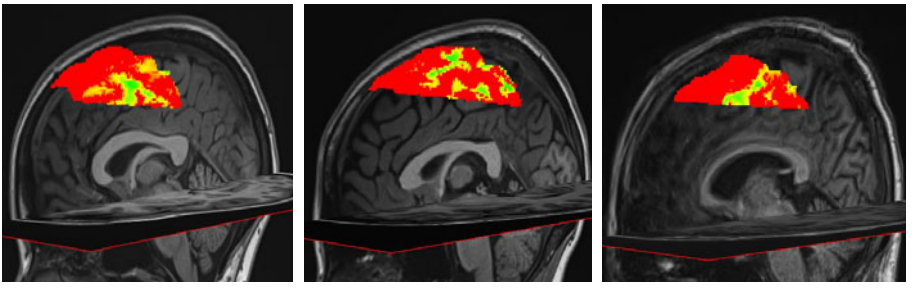
First, the senior resident was asked to perform manual trajectory planning of all 12 targets using the commercial Medtronic StealthStation™ software. The same 12 targets were passed to our automatic trajectory planning system. Table 1 shows a quantitative comparison between the risk, on a [1-100] scale, of manual and automatic trajectories with respect to each surgical constraint to remove any bias due to the choice of weights at the final aggregation stage. From Table 1, it is clear that automatically computed trajectory almost always exhibits a smaller risk for all individual criteria of ventricle, sulci and vessel avoidance.

Table 1. Comparison of automatic and manual planning (in parenthesis) for two normal subjects and one severe PD patient. The individual score for each surgical constraint is shown on a [1-100] scale. The automatic trajectory corresponds to the one with the lowest aggregated score and the manual trajectory was found by a senior neurosurgery resident using the Medtronic StealthStation™ platform.

	Target	Ventricles	Sulci		SWI vessels		TOF vessels	
		$risk_{max}$	$risk_{max}$	$risk_{sum}$	$risk_{max}$	$risk_{sum}$	$risk_{max}$	$risk_{sum}$
subj. 0	STN_{left}	1 (15)	2 (4)	2 (34)	1 (1)	7 (25)	3 (1)	10 (15)
	STN_{right}	1 (1)	1 (7)	2 (18)	4 (9)	16 (26)	2 (1)	23 (22)
	GPi_{left}	1 (22)	1 (8)	5 (27)	1 (16)	31 (56)	1 (1)	5 (8)
	GPi_{right}	1 (1)	2 (15)	7 (15)	1 (1)	5 (33)	1 (1)	2 (18)
subj. 1	STN_{left}	1 (31)	7 (22)	7 (57)	8 (9)	7 (27)	1 (1)	12 (49)
	STN_{right}	1 (1)	2 (45)	7 (43)	2 (39)	2 (93)	1 (1)	20 (45)
	GPi_{left}	1 (44)	2 (26)	7 (44)	1 (30)	32 (36)	1 (2)	32 (44)
	GPi_{right}	1 (11)	5 (24)	14 (40)	5 (12)	6 (51)	1 (1)	12 (39)
subj. PD	STN_{left}	1 (1)	5 (46)	13 (42)	4 (16)	37 (41)	2 (1)	5 (28)
	STN_{right}	1 (21)	5 (4)	21 (9)	4 (23)	15 (52)	1 (2)	3 (66)
	GPi_{left}	1 (1)	3 (7)	10 (24)	1 (1)	24 (54)	1 (2)	5 (35)
	GPi_{right}	12 (11)	1 (33)	3 (29)	4 (30)	5 (79)	1 (11)	12 (5)

Second, both surgeons were asked to qualitatively evaluate the 12 automatic trajectories selected by our software. This was done using 3D visualization software we developed and also by supplying the trajectories to the Medtronic StealthStation software because surgeons were most familiar with this platform. The surgeons found our automatic method more effective than manual planning for simultaneous avoidance of all critical structures. For examples: i) the manual trajectories to subject 0's STN_{left} and GPi_{left} travel near the ventral horn of the lateral ventricles; ii) because of subject 1's complex gyri pattern, manual sulci avoidance was sometime at a distance <1.5mm (instead of >2.5 mm with automatic planning), iii) the PD case exhibited larger ventricles and some brain atrophy, making the manual search more challenging.

Interestingly, the surgeons were able to identify new problems that are not handled by the current set of implemented constraints. For few suggested trajectories, they argued they would prefer a trajectory that is either more medial or that can avoid other basal ganglia nuclei (caudate, putamen). Only one suggested trajectory was found unsuitable because it was running parallel to a sulcus, at a safe distance, but in a way where a longer stretch of frontal lobe's cortex (grey matter) was traversed which could increase the risk of causing new neurological deficits.



Subject 0 Subject 1 PD patient

Fig. 3. Examples of color-coded trajectory map generated by our software. Color scale: green=recommended, yellow=acceptable, red=rejected.

4 Discussion and Conclusions

From the quantitative analysis of Table 1, it is clear that automatic planning outperforms manual planning with the task of aggregating several, well-defined, surgical constraints. Furthermore, automatic planning allows for thousands of trajectories (~12000) to be thoroughly analyzed in less than 4 minutes whereas manual planning approximately requires over 40 minutes per target.

Our experimental results showed that this quantitative analysis alone does not assess whether the implemented constraints mimic the complete decision-making process of neurosurgeons. Indeed, a qualitative analysis revealed the presence of previously undescribed surgical constraints. Fortunately, when an issue was raised an alternate entry point could easily be selected using the intuitive color-coded map outputted by our framework (see Fig. 3). In addition, we were able to eliminate most of these exceptional situations by adapting some of the user parameters. For example, we can force our software to find a more medial entry point simply by reducing the initial search to a more medial ROI.

In conclusion, manual path planning, especially with dense multi-modal datasets, is a complex and lengthy process that yields subjective and potentially sub-optimal solutions. This work provides neurosurgeons with an intuitive decision support system for automatic planning of DBS neurosurgeries that aggregates multiple surgical requirements into a single weighted ranking of available trajectories. Future work will concentrate on optimizing the weights attributed to each surgical constraint and on integrating additional constraints that were found to interplay with the decision-making process of surgeons to avoid long stretches of grey matter cortex and other basal ganglia nuclei (caudate and putamen).

Acknowledgement. This research was funded by the Natural Sciences and Engineering Research Council of Canada and the Centres of Excellence for Commercialization and Research.

References

1. de Lau, L.M.L., Breteler, M.M.B.: Epidemiology of Parkinson's disease. The Lancet Neurology 5, 525–535 (2006)
2. Chou, K.L.: Indication for Subthalamic Nucleus Deep Brain Stimulation Surgery. In: Baltuch, G.H., Stern, M.B. (eds.) Deep Brain Stimulation for Parkinson's Disease, pp. 41–54. informa healthcare, New York (2007)
3. Lee, J.D., Huang, C.H., Lee, S.T.: Improving stereotactic surgery using 3-D reconstruction. IEEE Eng. Med. Biol. Mag. 21, 109–116 (2002)
4. Nowinski, W.L., Yang, G.L., Yeo, T.T.: Computer-aided stereotactic functional neurosurgery enhanced by the use of the multiple brain atlas database. IEEE Trans. Med. Imaging 19, 62–69 (2000)
5. Fujii, T., Emoto, H., Sugou, N., Mito, T., Shibata, I.: Neuropath planner-automatic path searching for neurosurgery. International Congress Series 1256, 587–596 (2003)
6. Tirelli, P., De Momi, E., Borghese, N.A., Ferrigno, G.: An intelligent atlas-based planning system for keyhole neurosurgery. International Journal of Computer Assisted Radiology and Surgery 4, S85–S91 (2009)
7. Vaillant, M., Davatzikos, C., Taylor, R., Bryan, R.: A path-planning algorithm for image-guided neurosurgery. In: Troccaz, J., Mösges, R., Grimson, W.E.L. (eds.) CVRMed-MRCAS 1997, CVRMed 1997, and MRCAS 1997. LNCS, vol. 1205, pp. 467–476. Springer, Heidelberg (1997)
8. Brunenberg, E.J.L., Vilanova, A., Visser-Vandewalle, V., Temel, Y., Ackermans, L., Platel, B., ter Haar Romeny, B.M.: Automatic Trajectory Planning for Deep Brain Stimulation: A Feasibility Study. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part I. LNCS, vol. 4791, pp. 584–592. Springer, Heidelberg (2007)
9. Shamir, R.R., Tamir, I., Dabool, E., Joskowicz, L., Shoshan, Y.: A Method for Planning Safe Trajectories in Image-Guided Keyhole Neurosurgery. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6363, pp. 457–464. Springer, Heidelberg (2010)
10. Essert, C., Haegelen, C., Jannin, P.: Automatic Computation of Electrodes Trajectory for Deep Brain Stimulation. In: Liao, H., Edwards, P.J., Pan, X., Fan, Y., Yang, G.-Z. (eds.) MIAR 2010. LNCS, vol. 6326, pp. 149–158. Springer, Heidelberg (2010)
11. Haacke, E.M., Xu, Y., Cheng, Y.C., Reichenbach, J.R.: Susceptibility weighted imaging (SWI). Magn. Reson. Med. 52, 612–618 (2004)
12. Haacke, E.M., Mittal, S., Wu, Z., Neelavalli, J., Cheng, Y.C.: Susceptibility-weighted imaging: technical aspects and clinical applications, part 1. AJNR Am J. Neuroradiol. 30, 19–30 (2009)
13. Graves, M.J.: Magnetic resonance angiography. Br J. Radiol. 70, 6–28 (1997)
14. Collins, D.L., Zijdenbos, A.P., Baaré, W.F.C., Evans, A.C.: ANIMAL+INSECT: Improved Cortical Structure Segmentation. In: Kuba, A., Sámal, M., Todd-Pokropek, A. (eds.) IPMI 1999. LNCS, vol. 1613, pp. 210–223. Springer, Heidelberg (1999)
15. 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)

Automated Planning of Computer Assisted Mosaic Arthroplasty

Jiro Inoue[1,3], Manuela Kunz[1,3], Mark B. Hurtig[4], Stephen D. Waldman[2,3], and A. James Stewart[1,3]

[1] School of Computing
[2] Department of Mechanical Engineering, Queen's University
[3] Human Mobility Research Centre, Kingston General Hospital
[4] Ontario Veterinary College, University of Guelph

Abstract. We describe and evaluate a computer algorithm that automatically develops a surgical plan for computer assisted mosaic arthroplasty, a technically demanding procedure in which a set of osteochondral plugs are transplanted from a non-load-bearing area of the joint to the site of a cartilage defect. We found that the algorithm produced plans that were at least as good as a human expert, had less variability, and took less time.

1 Introduction

Cartilage degeneration is a widespread problem which occurs predominantly in the knee, ankle, and shoulder. Articular cartilage is not vascularized and does not naturally repair when damaged [1]. In the event of traumatic injury, this can lead to long-term damage and a loss of bearing capacity in the joint surface.

Mosaic arthroplasty is an operation that repairs damaged cartilage by transplanting osteochondral plugs from a non-load-bearing part of the joint to the site of the cartilage defect [2,3].

Mosaic arthroplasty is a technically challenging operation. The osteochondral plugs must be delivered so as to exactly reconstruct the original surface. But the top surface of a plug is curved and may not be perpendicular to the axis of the plug, so the position *and orientation* of each plug must be planned and achieved.

Plugs that are too high can result in poor plug integration due to micromotion and increased contact pressure [4,5]. Plugs that are too low can result in cartilage necrosis and fibrocartilage overgrowth [6]. A plug with an angled surface that is incorrectly delivered may be too high in some parts and too low in other parts. The percentage of the repair surface that is too high is inversely correlated with the quality of healing [7].

Computer assisted mosaic arthroplasty (CAMA) uses computer planning and intraoperative guidance to overcome these problems. The surgical plan indicates the position and orientation of each plug's harvest location and delivery location. The surgeon follows the plan using optically tracked surgical instruments or patient-specific instrument guides [8].

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 267–274, 2011.

The surgical plan is critical to the success of the operation. In the surgical plan:

- the original articular surface must be predicted as a reconstruction goal;
- the reconstructed cartilage thickness must match the original cartilage thickness to avoid degeneration at the bone/cartilage interface;
- the plug harvest sites must be planned so that surface orientation of each plug matches the predicted surface orientation at the corresponding delivery site; and
- an optimal pattern of plugs must be planned to cover the defect site with minimal overlapping and minimal gapping.

Manual planning for CAMA has been shown to be effective at finding harvest and delivery sites to reconstruct the original articular surface [9]. But manual planning is slow and requires a highly skilled human operator with knowledge of mosaic arthroplasty and the ability to use a 3D computer interface. The quality of a manually developed plan depends upon the skill of the operator and will be subject to inter-operator variability.

We describe a computer algorithm that automatically builds a surgical plan for CAMA. We show that the automatic planner achieves results that are at least as good as those of a skilled human operator and can reconstruct the correct surface to an accuracy of 0.3 mm RMS.

2 Related Work

Clinical evaluations have shown that surface congruency is critical to achieving a good repair. In an animal study [4], plugs that were too high were found to subside under weight bearing but showed fissuring, bone cysts, and poor bony incorporation. Elevated angled grafts (with an angled tip of the plug surface above the surrounding surface) are subject to increased contact pressure [5] which leads to peak loading and abrasion of the plug's surface [10]. On the other hand, plugs that are too low may be disposed to late degradation from inappropriate pressure [10] and have shown cartilage necrosis and fibrocartilage overgrowth [6].

Computer navigation of surgical tools has been shown to achieve greater plug perpendicularity and more appropriate plug depth [11] than with a freehand operation.

Computer planning and navigation have successfully been used in retrograde plug delivery in the human talus [12]. The planning consisted of using patient images to manually choose a path for the surgical drill. Another study using the talus found better plug harvesting and placement with computer assist [13]. Computer planning has also been used to match plugs from the femoral condyles to defects in the talus [14].

Computer imaging has been used to match cartilage topography at harvest and defect sites. One study determined the congruity between the surfaces at a harvest site and a donor site by finding the rigid transformation between

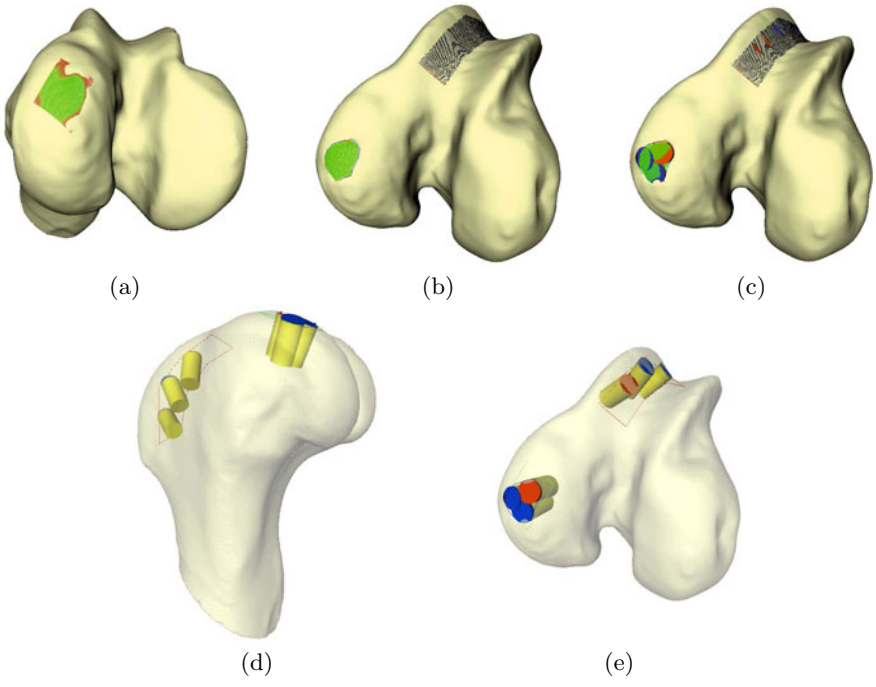


Fig. 1. The automatic planning process. (a) A bicubic spline is placed over the defect. (b) The defect area and harvest areas are outlined. (c) A pattern of plugs is made over the defect, and corresponding harvest sites found. (d,e) The final plan.

the surfaces that minimized the RMS error. The study applied this measure to seven cadaveric knees to propose the best harvest sites for particular defect sites [15]. Another study on six dog femurs used similar topographic matching and considered different plug diameters, finding that larger plug diameters had fewer good harvest sites [16].

Computer planning requires that the original articular surface be estimated, since it is no longer present at the site of the defect. A cubic spline can be used to accurately predict the original articular surface [17].

3 The Planning Algorithm

We found that a surprising simple planning algorithm could obtain results at least as good as those of a human operator.

The input to the algorithm consists of two coregistered triangle meshes: one of the cartilage surface and one of the underlying bone. These meshes can be obtained from segmented MRI or CT-arthrogram images. The output of the algorithm consists of the harvest and implant positions and orientations of a set of plugs, along with the plug dimensions.

In a preprocessing step, the original articular surface is estimated with a bicubic spline: The operator uses a 3D interface to pick four points on the cartilage surface around the defect and the algorithm creates a bicubic spline over the defect that interpolates those points and has the same tangent plan at those points (Figure 1(a)). This method has been shown to produce a surface within 0.27 mm RMS of the original, pre-defect surface [17].

Next, the operator outlines the defect on the spline surface. This is not done automatically because the defective cartilage usually extends beyond the area of visibly depressed cartilage (Figure 1(b)).

Next, the operator outlines areas of the joint from which cartilage plugs may be harvested (Figure 1(b)). This could be done automatically (but currently is not) by incorporating known good donor sites [15,16].

Finally, the algorithm is executed. The first phase of the algorithm builds an optimal pattern of plugs at the defect site. The second phase finds optimal harvest locations for each of the plugs. Results are shown in Figure 1(c,d,e).

Phase 1: Build plugs over defect

A hexagon grid is placed on the spline surface over the defect area. The initial position and orientation of the grid are arbitrary. Hexagon centers are spaced $\sqrt{3}\,r$ apart for plugs of radius r (in our case, $r = 2.25$ mm). A plug is placed at each hexagon center, oriented perpendicular to the spline surface.

The pattern of plugs should cover as much of the defect as possible while minimizing the overlapping of plugs, since plug stability may be reduced by overlapping [18]. To achieve this, the algorithm finds a pattern of plugs to minimize the cost function

$$A_{\text{uncovered}} + 0.4 \sum_i A_i$$

where $A_{\text{uncovered}}$ is the area of the defect not covered by any plug and A_i is the area of the i^{th} plug. An area covered by multiple plugs is counted multiple times in the sum, which discourages overlapping, but not at the expense of too much uncovered area. The constant 0.4 was chosen through experimentation to balance the number of plugs against the uncovered area.

For a given pattern of plugs, the cost function is approximated by sampling the defect area on a 0.1 mm square grid. Each grid point that is not covered by a plug counts $(0.1\text{ mm})^2$ and each plug of radius r counts $0.4\,\pi\,r^2$.

Ten thousand iterations of simulated annealing are performed to minimize the cost function. In each iteration, three cases are considered: add a random plug; remove a random plug; and keep the same plugs. In each of the three cases, the position of every plug is perturbed uniformly randomly within a $kr \times kr$ square centered at the plug's current position on the spline surface, where k is the "annealing temperature" which starts at one and decreases linearly with each iteration until it becomes zero in the last iteration. Of the three patterns considered, only the pattern of minimum cost is carried to the next iteration.

Note that plugs over the defect site are placed perpendicular to the surface. The algorithm could be extended to permit slanted plugs and to penalize subsurface plug intersection, in which a plug may be undercut by another plug,

making the undercut plug more likely to loosen. We also consider plugs of only one radius, although it would be easy for the algorithm to choose random radii from the set of radii available among the surgical tools.

Phase 2: Find harvest locations

Given a pattern of plugs over the defect, the second phase finds optimal locations at which to harvest those plugs. This is done in a greedy fashion: The plugs are randomly ordered and an optimal harvest location is found for each plug, in order.

Since the order affects the harvest locations (i.e. plugs that are later in the order have fewer locations from which to choose) the algorithm picks ten random orderings of the plugs, tries each, and chooses the ordering that results in the best harvest locations.

To find the best harvest location for a particular plug, a 0.3 mm square grid is placed over the harvest areas and each grid point is considered as a harvest location. (Grid points that are within one diameter of an already-harvested plug are discarded because the joint is weakened when harvest holes are too close together.) For a particular grid point, the plug from the defect site is translated so that the middle of its top surface is coincident with the grid point and 49 orientations are considered by varying the angle from perpendicular in 5 degree increments from -15 to +15 degrees in both the x and y directions of the grid. At each orientation, the rotation around the plug axis and the translation along the plug axis are determined that give the best fit between the surface of the plug and the surface at the harvest site. The best fitting plug is chosen and its RMS surface error is used as the cost.

A variant of the algorithm tries to match both the cartilage surface and the bone surface. In this case, 49 orientations are again considered and the best axial rotation and translation are found for the cartilage surface and, separately, for the bone surface. Then each angle between the two axial rotations is tested in 0.5 degree increments. At each such angle, the best-fit translation along the axis is found. In this variant, the cost of a harvest location is the sum of the RMS errors for the two surfaces.

4 Evaluation of the Algorithm

The algorithm was tested on twelve in-vivo sheep knees for which data was gathered as part of a larger study. Each knee was scanned using a CT arthrogram with a slice thickness of 0.625 mm. The bone and cartilage surfaces were manually segmented from the CT images. In a minimally invasive surgery, a cartilage defect was induced on the medial condyle with a calibrated impact. Three months later, a second CT arthrogram was taken and the bone and cartilage surfaces manually segmented.

The post-defect models were used for planning and the resulting plans were evaluated in comparison to the pre-defect models. Three conditions were tested for each knee:

- An expert human operator, who had substantial experience planning mosaic arthroplasty, manually produced a surgical plan. The operator matched only the cartilage surface and did not attempt to match the bone surface.
- An inexperienced operator used the planning algorithm to produce a surgical plan that optimized the fit of only the cartilage surface.
- The inexperienced operator used the planning algorithm to produce a surgical plan that optimized the fit of both the cartilage surface and the bone surface.

Each plan was evaluated for surface congruency by computing the RMS error between the planned cartilage surface and the pre-defect cartilage surface of the pre-defect CT arthrogram. The RMS calculation used six points on the circumference and one point at the center of each plug surface. A similar RMS calculation was made for congruency at the bone surface. The percentage of the defect surface covered by plugs was calculated. The total planning time was recorded.

5 Results

Table 1 summarizes the results. For manual and automatic planning that considered only the cartilage surface (evaluating significance with one-sided t-tests):

- No significant difference was found in the RMS errors (manual 0.31 mm, automatic 0.25 mm, $p = 0.085$), although the low p value is suggestive of a trend in favor of automatic planning.
- Automatic planning shows a tighter variance in RMS error ($\pm$ 0.06, min 0.13, max 0.38) than manual planning ($\pm$0.10, min 0.09, max 0.61), but not to a statistically significant degree using Levene's test for equal variances ($p = 0.17$).
- No significant difference was found in the coverage of the defect surface (manual 84%, automatic 88%, $p = 0.10$), although the trend is also in favor of automatic planning.
- There was significantly less plug overlap with automatic planning (manual 16.1%, automatic 9.7%, $p = 0.02$).
- The automatic planning time was substantially faster (4.5 minutes) than the reported manual planning time (30 to 45 minutes), but we cannot make any statistical claims because the manual planning time was not measured accurately.

When the automatic planner optimized both the cartilage and bone surfaces, the defect coverage remained the same at 88% and the total planning time remained the same at 4.5 minutes. The RMS error of the cartilage surface increased, as one would expect when the bone surface also had to be considered in the optimization. Interestingly, the overall (bone plus cartilage) RMS error was *less* when only cartilage was considered (although not by a statistically significant amount, so we cannot attribute any meaning to this).

Table 1. Results showing, for each measure, "average ± 95% confidence interval". In the *Condition* column, "C" denotes cartilage only and "C+B" denotes cartilage and bone. * Unfortunately, the manual planning was not timed and the 30 to 45 minutes shown in the table is an estimate provided by the operator. Automatic planning time includes the time for the operator to set up the spline and regions.

Condition	RMSE cartilage (mm)	RMSE cartilage and bone (mm)	Coverage (%)	Overlap (%)	Time (minutes)
Manual, C	0.31 ± 0.10	0.31 ± 0.07	84 ± 7	16.1 ± 4.9	$30 - 45^*$
Auto, C	0.26 ± 0.06	0.27 ± 0.07	88 ± 3	9.7 ± 4.2	4.5 ± 1.7
Auto, C+B	0.35 ± 0.09	0.29 ± 0.06	88 ± 3	9.7 ± 4.2	4.5 ± 1.9

6 Conclusions

We have described a computer algorithm to automatically develop a surgical plan for computer assisted mosaic arthroplasty. No statistically significant difference was found between the RMS error of automatic planning and manual planning, although the trends suggest that automatic planning gives better coverage, is more accurate, and has less variance. Automatic planning produces less plug overlap (a good feature) and the algorithm takes much less time than a highly trained human expert and can be used by an inexperienced operator. These results suggest that the automatic planner should be used in place of manual planning, with the plans being subject to final approval from the surgeon.

In ongoing work, we are augmenting the automatic planner with multiple plug radii and non-perpendicular plugs at the defect site, and will shortly be using the automatic planner in human surgeries.

Acknowledgements. Thank you to the reviewers for their many helpful comments. This work is supported by Natural Sciences and Engineering Research Council of Canada (NSERC) grant STPGP 336797 and by the joint NSERC/ Canadian Institutes of Health Research CHRP grant 385959.

References

1. Halser, E., Herzog, W., Wu, J.Z., Müller, W., Wyss, U.: Articular cartilage biomechanics: theoretical models, material properties, and biosynthetic response. Crit. Rev. Biomed. Eng. 27(6), 415 (1999)
2. Yamashita, F.: The transplantation of an autogeneic osteochondral fragment for osteochondritis dissecans of the knee. Clin. Orthop. Relat. Res. (201), 43 (1985)
3. Hangody, L., Kish, G., Krpti, Z., Szerb, I., Udvarhelyi, I.: Arthroscopic autogenous osteochondral mosaicplasty for the treatment of femoral condylar articular defects: A preliminary report. Knee Surgery, Sports Traumatology, Arthroscopy 5(4), 262–267
4. Pearce, S., Hurtig, M., Clarnette, R., Kalra, M., Cowan, B., Miniaci, A.: An investigation of 2 techniques for optimizing joint surface congruency using multiple cylindrical osteochondral autografts. Arthroscopy: The Journal of Arthroscopic and Related Surgery 17(1), 50–55 (2001)

5. Koh, J., Kowalski, A., Lautenschlager, E.: The effect of angled osteochondral grafting on contact pressure: A biomechanical study. American Journal of Sports Medicine 34(1), 116–119 (2006)
6. Huang, F., Simonian, P., Norman, A., Clark, J.: Effects of small incongruities in a sheep model of osteochondral autografting. American Journal of Sports Medicine 32(8), 1842–1848 (2004)
7. Kunz, M., Hurtig, M., Waldman, S., Devlin, S., Rudan, J., Bardana, D., Stewart, J.: Image-guided surgical techniques for cartilage repair - an animal trial. In: World Congress of the International Cartilage Repair Society (ICRS), p. 190 (2010)
8. Radermacher, K., Portheine, F., Anton, M., Zimolong, A., Kaspers, G., Rau, G., Staudte, H.W.: Computer assisted orthopaedic surgery with image based individual templates. Clin. Orthop. Relat. Res. (354), 28–38 (1998)
9. Kunz, M., Devlin, S., Rudan, J., Waldman, S., Stewart, J., Hurtig, M.: Computer-assisted planning for mosaic arthroplasty. Computer Assisted Radiology and Surgery, S102–S103 (2009)
10. Jakob, R., Franz, T., Gautier, E., Mainil-Varlet, P.: Autologous osteochondral grafting in the knee: indication, results, and reflections. Clinical Orthopaedics and Related Research 401, 170–184 (2002)
11. Koulalis, D., Benedetto, P.D., Citak, M., Loughlin, P., Pearle, A., Kendoff, D.: Comparative study of navigated versus freehand osteochondral graft transplantation of the knee. American Journal of Sports Medicine 37(4), 803–807 (2009)
12. Bale, R., Hoser, C., Rosenberger, R., Rieger, M., Benedetto, K., Fink, C.: Osteochondral lesions of the talus: Computer-assisted retrograde drilling - feasibility and accuracy in initial experiences. Radiology 218, 278 (2001)
13. Hoser, C., Bichler, O., Bale, R., Rosenberger, R., Rieger, M., Kovacs, P., Lang, T., Fink, C.: A computer assisted surgical technique for retrograde autologous osteochondral grafting in talar osteochondritis dissecans (ocd): a cadaveric study. Knee Surgery, Sports Traumatology, Arthroscopy 12(1), 65–71 (2004)
14. Marymount, J., Shute, G., Zhu, H., Varner, K., Paravic, V., Haddad, J., Noble, P.: Computerized matching of autologous femoral grafts for the treatment of medial talar osteochondral defects. Foot and Ankle Intl. 26(9), 708 (2005)
15. Bartz, R., Kamaric, E., Noble, P., Lintner, D., Bocell, J.: Topographic matching of selected donor and recipient sites for osteochondral autografting of the articular surface of the femoral condyles. American Journal of Sports Medicine 29(2), 207–212 (2001)
16. Böttcher, P., Zeissler, M., Grevel, V., Oechtering, G.: Computer simulation of the distal aspect of the femur for assessment of donor core size and surface curvature for autologous osteochondral transplantation in the canine stifle joint. Veterinary Surgery 39(3), 371–379 (2010)
17. Kunz, M., Devlin, S., Gong, R.H., Inoue, J., Waldman, S.D., Hurtig, M., Abolmaesumi, P., Stewart, J.: Prediction of the repair surface over cartilage defects: A comparison of three methods in a sheep model. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5761, pp. 75–82. Springer, Heidelberg (2009)
18. Haklar, U., Tuzuner, T., Uygur, I., Kocaoglu, B., Guven, O.: The effect of overlapping on the primary stability of osteochondral grafts in mosaicplasty. Knee Surgery, Sports Traumatology, Arthroscopy 16(7), 651–654 (2008)

Model-Updated Image-Guided Minimally Invasive Off-Pump Transcatheter Aortic Valve Implantation

Mohamed Esmail Karar[1], Matthias John[2], David Holzhey[3], Volkmar Falk[4], Friedrich-Wilhelm Mohr[1,3], and Oliver Burgert[1]

[1] Innovation Center Computer Assisted Surgery (ICCAS), University of Leipzig, Germany
[2] Siemens AG, Healthcare Sector, Forchheim, Germany
[3] Department of Cardiac Surgery, Heart Center, University of Leipzig, Germany
[4] Division of Heart and Vascular Surgery, University Hospital Zurich, Switzerland
mohamed.karar@medizin.uni-leipzig.de

Abstract. This paper presents a method for assisting the placement of stented aortic valve prosthesis during minimally invasive off-pump transcatheter aortic valve implantation (TAVI) under live 2-D X-ray fluoroscopy guidance. The proposed method includes a dynamic overlay of an intra-operative 3-D aortic root mesh model and an estimated target area of valve implantation onto live 2-D fluoroscopic images. This is based on a template-based tracking procedure of a pigtail catheter without further injections of contrast agent. Minimal user-interaction is required to initialize the algorithm and to correct fluoroscopy overlay errors if needed. Retrospective experiments were carried out on ten patient datasets from the clinical routine of the TAVI. The mean displacement errors of the updated aortic root mesh model overlays are less than 2.0 mm without manual overlay corrections. The results show that the guidance performance of live 2-D fluoroscopy is potentially improved when using our proposed method for the TAVIs.

Keywords: Transcatheter aortic valve implantation, X-ray fluoroscopy, image-guided interventions, aortic valve prosthesis.

1 Introduction

Aortic valve replacement is the standard treatment of degenerative aortic valve stenosis. However, many elderly patients have an elevated predicted operative risk that could compromise the patient's outcome after standard open heart surgery [1]. Transcatheter aortic valve implantation (TAVI) therefore presents a good alternative to the standard surgical treatment for elderly and high-risk patients with severe aortic stenosis [1].

The transapical TAVI is a recent and minimally invasive, off-pump technique that consists of a left anterolateral mini-thoracotomy for direct antegrade surgical access through the apex of the left ventricle [2]. This is followed by the insertion of an inflatable bioprosthetic valve through a catheter and the implantation within the diseased, native aortic valve under rapid ventricular pacing (RVP). During the intervention, the placement of stented aortic valve prosthesis such as the Edwards

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 275–282, 2011.

Sapien™ THV stent-valve (Edwards Lifesciences, Irvine, USA) is crucial and is typically performed under two-dimensional (2-D) X-ray fluoroscopy guidance [1-3].

Live 2-D fluoroscopy displays 2-D projection images of interventional devices such as intra-cardiac catheters, aortic valve prosthesis (AVP) and some anatomical structures. Cardiac blood vessels, such as the aortic root and coronary arteries, are only visible in single-plane fluoroscopic images if the contrast agent is injected using a pigtail catheter for guiding the TAVI [2]. However, the dosages of contrast injections must be minimized because of renal insufficiencies in elderly patients. Before starting the TAVI procedure, the surgeon can use an angiographic C-arm system to reconstruct intra-operative three-dimensional (3-D) CT-like images of the aortic root under RVP from an acquired rotational 2-D fluoroscopic image sequence with 15-25 ml contrast agent of 200^{o} over 5 seconds [3], see Fig. 1. This small amount of contrast agent is equivalent to just a single angiogram aortic root shot.

There are few previous studies related to image-guided TAVI interventions. In [4], an intra-operative guidance system has been proposed to include the planning system of [5] and to perform real-time tracking of the AVP in fluoroscopic images. A system for automatic segmentation and for the static overlay of aortic root volume and landmarks on live fluoroscopic images is described in [6,7]. Using intra-operative MRI guidance, a new robotic assistance system has been evaluated for delivering the AVP using a phantom [8].

In this paper, we aim to automatically display an appropriate placement position of the transcatheter AVP using live 2-D fluoroscopy guidance. Based on the motion of a pigtail catheter in the absence of contrast agent, the proposed method approximates the motion of the aortic root to update the overlay of a 3-D aortic root mesh model including anatomical valve landmarks such as coronary ostia from intra-operative 3-D C-arm CT images onto 2-D fluoroscopic images. If the contrast agent appears in fluoroscopic images, this fluoroscopy overlay is not required and can be switched off [9]. Furthermore, automatic estimation of a target area of implantation is visualized to potentially improve the accuracy of the implanted AVP.

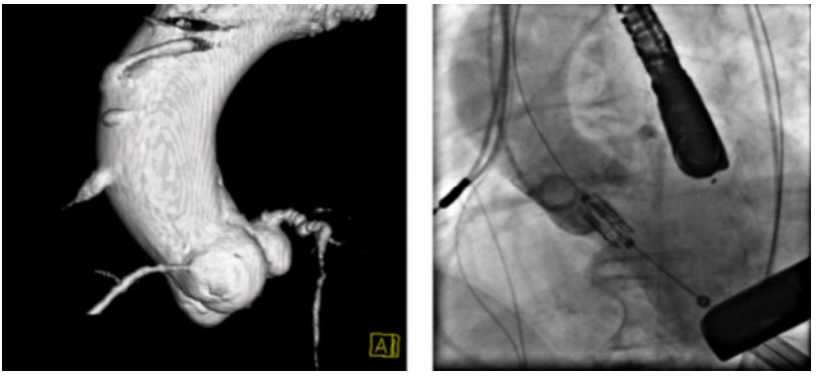


Fig. 1. Intra-operative 3-D C-arm CT image of the aortic root (*left*) and 2-D fluoroscopy guidance for assisting the placement of aortic valve prosthesis (*right*)

2 Method

The proposed method starts with algorithm initialization, followed by an updating procedure of the overlaid aortic root model. In the initialization step, a 3-D geometrical mesh model of the aortic root with valve landmarks is acquired from the interventional C-arm system. Based on the best experience and knowledge of the physician, a target area of valve implantation is automatically estimated inside the 3-D mesh model. The aortic root model and the target area of valve implantation are then projected and manually aligned onto a contrast image. The contrast image is a fluoroscopic image which is automatically detected to display the aortic root roadmap with contrast agent. Finally, the overlay of the aortic root model and the target area of implantation onto live fluoroscopic images are updated by following the aortic root motion via the tracking of a pigtail catheter.

In the following, each component of this proposed method is described in more detail.

2.1 Aortic Root Mesh Model and Target Area of Implantation

In order to guide the TAVI procedure, a 3-D triangulated mesh model of the aortic root has been used. The mesh geometry is generated based on a learning-based 3-D boundary detector of the aortic root in intra-operative C-arm CT images [6]. Using a discriminative learning-based landmark detector, the resulting model moreover includes eight anatomical valve landmarks [6]; namely two points of coronary ostia (left LCO_{3d} and right RCO_{3d}), three points of commissures (left LC_{3d}, right RC_{3d} and non-coronary NC_{3d}) and the three lowest points (hinge points) of each leaflet cusp (left LLC_{3d}, right RLC_{3d} and non-coronary NLC_{3d}), see Fig. 2 (*left*) .

Based on professional surgical experience, the correct placement of AVP should be one-third to one-half of its length above and perpendicular to the aortic annulus [2]. In this study, a target area of valve implantation is defined by two embedded circles of annulus and ostia planes with the normal center line to the valve annulus [5]. The annulus circle $AC_{3d} = C\ (c_a,\ r_a)$ is defined by the three lowest points of the valve cusps. The circle's center c_a is the centroid of the cusp points (LLC_{3d}, RLC_{3d}, NLC_{3d}). The radius r_a is calculated from the lengths of the sides of the cusp points (a, b, c) and the area of three cusps points A_a from heron's formula:

$$r_a = \frac{abc}{4A_a}, \text{ where } A_a = \sqrt{\frac{(a^2+b^2+c^2)^2 - 2(a^4+b^4+c^4)}{4}}. \tag{1}$$

The circle is inscribed to the corresponding plane of cusp points through c_a normal to

$$n_a = \frac{(c_a - LLC_{3d}) \times (c_a - RLC_{3d})}{\|c_a - LLC_{3d}\| \cdot \|c_a - RLC_{3d}\|}. \tag{2}$$

Similarly, the ostia circle OC_{3d} is calculated at the level of the lowest coronary ostium within the aortic root and parallel to the AC_{3d}. The normal center line CL_{3d} is connected between the two centers of AC_{3d} and OC_{3d} as shown in Fig. 2 (*left*).

2.2 Alignment of Aortic Root Model onto Contrast Image

By analyzing the histogram of live fluoroscopic images and using the 98-percentile as a threshold measurement of contrast agent, a contrast image is automatically detected after learning the histogram feature curve of the first 20 frames of the fluoroscopic sequence without contrast agent [9]. In this contrast image, the aortic root roadmap shows up in dark pixels as depicted in Fig. 2 (*middle*).

The 3-D aortic mesh model with landmarks and the target area of implantation are then projected onto a 2-D fluoroscopy plane by using the transformation matrix of the interventional C-arm system. This projected data is initially aligned with the aortic root roadmap onto the contrast image by the physician, as shown in Fig. 2 (*right*).

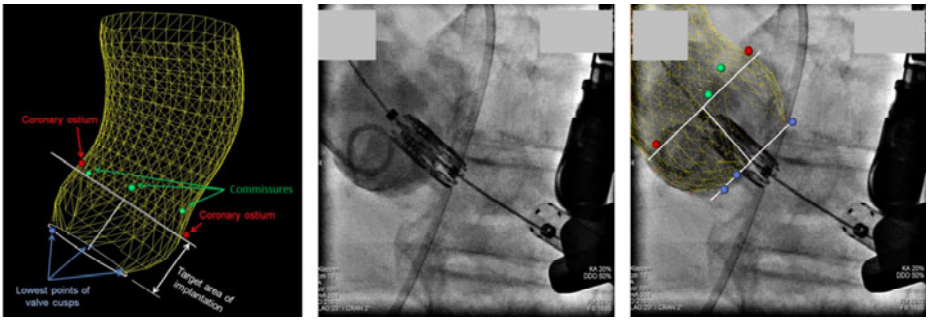


Fig. 2. 3-D aortic root mesh model including anatomical valve landmarks and target area of implantation (*left*), a contrast image shows up the aortic root roadmap with contrast agent (*middle*), and the manual alignment of projected aortic mesh model and target area of implantation with the aortic roadmap onto the contrast image (*right*)

2.3 Updating Aortic Root Model Overlay

The collaborating physicians confirmed that the aortic root moves with the pigtail catheter unless it is pulled or pushed manually. Therefore, we assumed the image-based tracking procedure of the pigtail catheter to provide a dynamic overlay of the projected aortic mesh model with landmarks and the target area of implantation onto live 2-D fluoroscopy as follows. Firstly, fluoroscopic images are preprocessed using a 2-D Gabor filter [10], in order to reduce the image noise and to enhance the features of the pigtail catheter. Secondly, a template matching approach [11] is applied to determine the current position of the pigtail catheter within each fluoroscopic image of the live sequence, see Fig. 3 (*left*). The template image of the pigtail catheter t is manually defined on the first image of the sequence. A region of interest (ROI) is defined to reduce the processing time and increase the algorithm robustness. In practice, the size of the ROI is chosen to be 2.5x the size of the template image.

In this matching approach, $I(x,y)$ denotes the intensity of a preprocessed ROI image of the size $S_x \times S_y$ at point (x,y), $x \in \{0, ..., S_x\text{-}1\}$, $y \in \{0, ..., S_y\text{-}1\}$ and the template image t of the size $s_x \times s_y$. The position of the catheter is determined by a pixelwise comparison of the ROI image with the template image, based on the computing of fast

normalized cross correlation (FNCC) coefficient γ at each point *(u, v)* for ROI and for template images. Eq. 3 provides the definition of γ. Here $\bar{i}_{u,v}$ and $\bar{t}$ are the mean brightness values within the ROI and the template image respectively. The normalized value γ_{max} at the point *(u, v)* in the current ROI image defines the best matching location of the template.

$$\gamma(u,v) = \frac{\sum_{x,y}(I(x,y) - \bar{i}_{u,v})(t(x-u, y-v) - \bar{t})}{\sqrt{\sum_{x,y}(I(x,y) - \bar{i}_{u,v})^2 \sum_{x,y}(t(x-u, y-v) - \bar{t})^2}}. \quad (3)$$

Finally, the translational motion of the visualized aortic root model, valve landmarks and the target area of implantation is then updated by calculating the displacement $\mathbf{d} = (d_x, d_y)$ of pigtail catheter between two images of the sequence [12]. This is based on the different matching positions of the catheter template between these two images. In order to ensure the high accuracy of the tracked pigtail catheter, the template-based tracking algorithm can be temporarily stopped if the best matching value of γ_{max} is less than 50% in the current proceeded image.

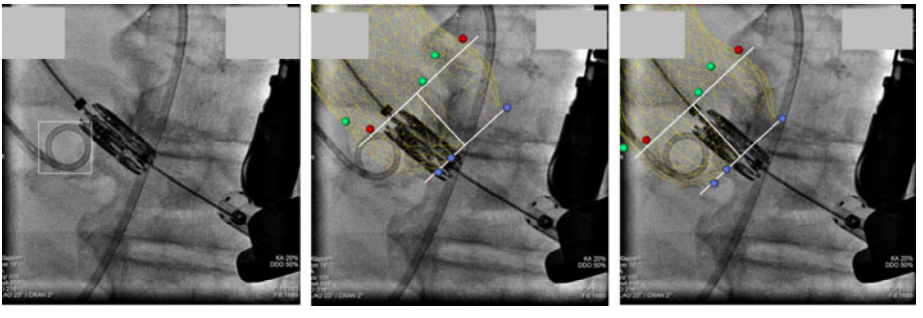


Fig. 3. Template matching of a pigtail catheter (*left*), projected aortic mesh model with landmarks and target area of implantation onto a fluoroscopic image of live sequence with static overlay (*middle*) and with updating model overlay (*right*)

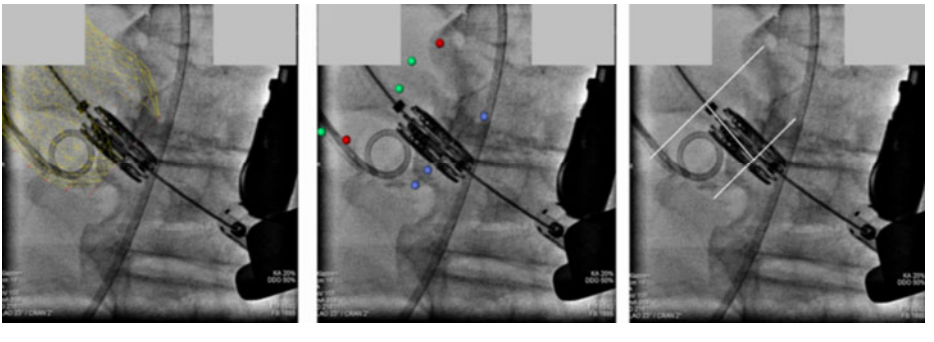


Fig. 4. Views of visualized aortic mesh model (*left*), valve landmarks (*middle*), and target area of implantation (*right*) projected onto a fluoroscopic image

2.4 User Interaction and Visualization

An interactive graphical user interface (GUI) is integrated within the proposed method based on C++ and open source components such as the Visualization Toolkit (VTK) [13]. The parameter values needed to compute the 3-D to 2-D transformation matrix of the C-arm imaging system are imported from a Digital Imaging and Communications in Medicine (DICOM) file or given by the user. Moreover, different views of the overlaid mesh model, valve landmarks and target area of implantation are separately visualized to allow the physician to display only the required information for the prosthesis deployment, as depicted in Fig. 4.

3 Results and Evaluation

In order to test and evaluate the proposed method, experiments were retrospectively carried out on different patient datasets from the clinical routine of the TAVI. These datasets included ten fluoroscopic image sequences and related aortic root mesh models of three female and two male patients around 84 years of age. Each fluoroscopic sequence included 90 images with 512 × 512 to 1024 × 1024 pixels. The pixel size was approximately 0.15 mm. All fluoroscopic images and aortic root models were acquired from the interventional C-arm system (Artis zeego, Siemens AG, Healthcare Sector, Forchheim, Germany) at the Heart Center Leipzig, Germany.

Using a standard PC with Intel CPU (2.4 GHz), the total computation times of the algorithm initialization and dynamic overlay procedures were approximately three to five minutes and 100 to 125 milliseconds per frame respectively.

The updating performance of overlaid aortic root model onto live fluoroscopic images was indirectly evaluated by calculating the absolute mean and maximum displacement errors, d_{mean} ± standard deviation (SD) and d_{max}, between the expected and the computed displacements of the pigtail template over all tested fluoroscopic images, see Fig. 5. The expected displacements of the tracked catheter are manually performed by an experienced user during the evaluation procedure only. The evaluation results did not include the images with high doses of contrast agent, because these contrasted images (three to ten images per each sequence) can temporarily stop the template-based tracking algorithm of the pigtail catheter.

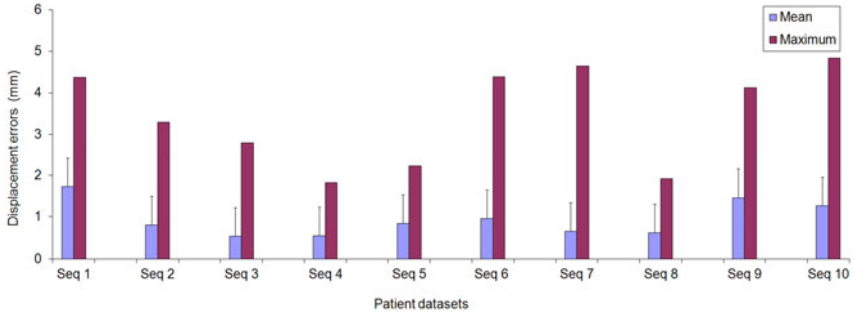


Fig. 5. Evaluation results of the updated aortic root model overlay onto fluoroscopic image sequences for ten patient datasets based on displacement errors of the pigtail catheter

As depicted in Fig. 5, the fluoroscopic image sequence (seq. 1) shows relatively high displacement errors of d_{mean} = 1.73 ± 0.86 mm and d_{max} = 4.37 mm. This is due to the pigtail catheter having been slightly pulled by the physician.

Due to an overlapping of the pigtail catheter with the prosthesis, image sequences 7 and 10 yielded the highest maximum displacement errors 4.65 mm and 4.84 mm, respectively, but have such large errors for only one to three images per sequence. However, all tested fluoroscopic images showed that the absolute mean displacement errors are less than 2.0 mm and within the clinically accepted ranges.

4 Discussion

We have proposed a method to assist the placement of the AVP during minimally invasive off-pump TAVI under 2-D fluoroscopy guidance. Compared to the work in [7], this proposed method can avoid misalignments between the static overlay of the aortic root model and invisible aortic root roadmaps in live fluoroscopic images during the TAVI procedure. Furthermore, automatic definition and visualization of the target area of implantation is potentially saving the physician's effort and surgery time as well.

In order to perform dynamic overlay of the aortic root model and the target area of implantation onto live fluoroscopic images, the template matching using FNCC approach has been applied for tracking the pigtail catheter which approximates the translational motion of the aortic root without contrast agent injections. Interventional devices such as the transesophageal echocardiography (TEE) probe and guide wires have no effect on the detection of the pigtail catheter, because they often lie outside of the ROI. Overlapping with the AVP may affect the matching accuracy of the pigtail catheter. Nevertheless, the obtained results demonstrated that our method is robust enough to track the pigtail catheter in the context of the TAVI procedure.

Unfortunately a clinical validation of this study would require additional contrast agent injections which are dangerous for elderly and high-risk patients who have renal insufficiencies. Because there are no tools or ground truth datasets to accurately identify the correct overly of the aortic root model without contrast injections, the updating accuracy of overlaid aortic root models has therefore been indirectly evaluated by estimating the displacement errors of the pigtail catheter for all ten patient datasets, see Fig. 5. The collaborating physicians assumed that the displacement errors for the TAVI were 2 to 5 mm, and that the margin of errors should not exceed 2 mm in narrow calcific aortic stenosis. In Fig. 5, the evaluation results show that the absolute mean errors are approximately less than 2.0 mm within the clinically accepted ranges. An interactive GUI has been developed to manually correct the high displacement errors by adding suitable offset values in X-Y directions for the tracked pigtail catheter.

Using this image-based method to guide the TAVI procedure, an assistance system in a "hybrid" operating room is under development which is connected to the interventional C-arm system and is capable of real-time clinical studies.

Acknowledgments. The Authors would like to thank Alois Nöttling and Stefan Ammon for the technical support during this study. This work is supported by German Academic Exchange Service (DAAD) under scholarship number A0690520. This work is partially funded by Siemens AG, Forchheim, Germany.

References

1. Ferrari, E., Sulzer, C., Marcucci, C., Qanadly, S.D., Locca, D., Berdajs, D., Von Segesser, L.K.: Transcatheter Aortic Valve Implantation (TAVI): State of The Art Techniques and Future Perspectives. Swiss. Med. Wkly. 140, w13127 (2010)
2. Walther, T., Dewey, T., Borger, M.A., Kempfert, J., Linke, A., Becht, R., Falk, V., Schuler, G., Mohr, F.W., Mack, M.: Transapical Aortic Valve Implantation: step by step. Ann. Thorac. Surg. 87, 276–283 (2009)
3. Kempfert, J., Falk, V., Schuler, G., Linke, A., Merk, D., Mohr, F.W., Walther, T.: Dyna-CT During Minimally Invasive Off-Pump Transapical Aortic Valve Implantation. Ann. Thorac. Surg. 88, 2041 (2009)
4. Karar, M.E., Gessat, M., Walther, T., Falk, V., Burgert, O.: Towards a New Image Guidance System for Assisting Transapical Minimally Invasive Aortic Valve Implantation. In: Conf. Proc. IEEE Eng. Med. Biol. Soc., pp. 3645–3648. IEEE Press, New York (2009)
5. Gessat, M., Merk, D.R., Falk, V., Walther, T., Jacobs, S., Nottling, A., Burgert, O.: A Planning System for Transapical Aortic Valve Implantation. In: Proc. of SPIE Medical Imaging (2009)
6. Zheng, Y., John, M., Liao, R., Boese, J., Kirschstein, U., Georgescu, B., Zhou, S.K., Kempfert, J., Walther, T., Brockmann, G., Comaniciu, D.: Automatic Aorta Segmentation and Valve Landmark Detection in C-Arm CT: Application to Aortic Valve Implantation. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6361, pp. 476–483. Springer, Heidelberg (2010)
7. John, M., Liao, R., Zheng, Y., Nottling, A., Boese, J., Kirschstein, U., Kempfert, J., Walther, T.: System to Guide Transcatheter Aortic Valve Implantations Based on Interventional C-arm CT Imaging. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6361, pp. 375–382. Springer, Heidelberg (2010)
8. Ming, L., Kapoor, A., Mazilu, D., Horvath, K.A.: Pneumatic Actuated Robotic Assistant System for Aortic Valve Replacement Under MRI Guidance. IEEE Trans. on Biomed. Eng. 58, 443–451 (2011)
9. Condurache, A.P., Aach, T., Eck, K., Bredno, J.: Fast Detection and Processing of Arbitrary Contrast Agent Injections in Coronary Angiography and Fluoroscopy. In: Proc. of Bildverarbeitung fuer die Medizin (2004)
10. Kong, W.K., Zhang, D., Li, W.: Palmprint Feature Extraction Using 2-D Gabor Filters. J. Pattern Recognition 36, 2339–2347 (2003)
11. Briechle, K., Hanebeck, U.D.: Template Matching Using Fast Normalized Cross Correlation. In: Proc. of SPIE Optical Pattern Recognition XII (2001)
12. Atasoya, S., Groher, M., Darko, Z., Glocker, B., Waggershauser, T., Pfister, M., Navab, N.: Real-Time Respiratory Motion Tracking: Roadmap Correction for Hepatic Artery Catheterizations. In: Proc. of SPIE Medical Imaging (2008)
13. Visualization Toolkit, Kitware Inc., `http://www.vtk.org`

Real-Time 3D Ultrasound Guided Interventional System for Cardiac Stem Cell Therapy with Motion Compensation

Vijay Parthasarathy[1], Charles Hatt[2], Zoran Stankovic[3], Amish Raval[2], and Ameet Jain[1]

[1] Philips Research North America, Briarcliff Manor, NY
[2] University of Wisconsin Medical School, Madison, WI
[3] Philips Research Eindhoven, The Netherlands

Abstract. This paper describes a clinically translatable interventional guidance platform to improve the accuracy and precision of stem cell injections into a beating heart. The proposed platform overlays live position of an injection catheter onto a fusion of a pre-procedural MR roadmap with real-time 3D transesophageal echocardiography (TEE). Electromagnetic (EM) tracking is used to initialize the fusion. The fusion is intra-operatively compensated for respiratory motion using a novel algorithm that uses peri-operative full volume ultrasound images. Validation of the system on a moving heart phantom produced a landmark registration accuracy of 2.8 $\pm$ 1.45mm. Validation on animal *in vivo* data produced an average registration accuracy of 2.2 $\pm$ 1.8 mm; indicating that it is feasible to reliably and robustly fuse the MR road-map with catheter position using 3D ultrasound in a clinical setting.

Keywords: Stem cell therapy, Motion compensation, 3D TEE, EM tracking, Image fusion, Interventional cardiology.

1 Introduction

Stem cell repair of recently infarcted tissue could be a potential cure for patients with recent heart attacks [1]. One way to deliver stem cells to the infarcted region of a heart is through direct myocardial injection using a catheter. These cardiac injections need to be precisely targeted in order to avoid puncturing the infarcted portion; thereby needing precise localization of the catheter with respect to the anatomy of the heart that is moving due to both cardiac and respiratory motion. In this work, we aim to provide accurate localization by integrating three coordinate systems: 1) pre-operative MR road map, 2) live-3D ultrasound (US) using transesophageal echo, and 3) injection catheter. Such an interventional fusion system can help visualize the live 3D ultrasound volumes and the catheter in the larger context of the pre-procedural planning volumes. For this solution to be clinically viable in interventional cardiology and electrophysiology procedures, the set of registrations required to achieve this fusion needs to be accurate, fast, and robust, *i.e.* to be able to maintain continuous real-time registration between the different coordinate systems in free breathing mode for the entire procedure.

Maintaining continuous registration requires continuous re-adjustment of registration due to cardiac and respiratory motion. Continuous registrations of preoperative

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 283–290, 2011.

volumetric images with real-time intraoperative ultrasound images are challenging and computationally demanding. In the case of real-time 3D ultrasound imaging probes like TEE and intracardiac echo (ICE), the accuracy and robustness of continuous registrations is compromised due to three factors: 1) small field-of-view (FOV) of ultrasound images, 2) artifacts and limited signal-to-noise ratio of *in vivo* ultrasound images, and 3) very different contrast mechanisms between pre-procedural images and ultrasound images.

Related Work. The demands of speed, accuracy, and robustness in procedures involving real-time registration have been addressed in several ways. Addressing the speed issue, Huang *et al.*, proposed using very small number of pixels along the edges of the image volumes for doing a mutual information based spatiotemporal registration [2]. This registration, however, was performed on data collected during breath-hold. Similarly, Sun et al [3], performed registration between 2D ICE and C-arm-CT volumes using cardiac and respiratory gating. To address the problem of different contrasts, King *et al.* [4] and Wein *et al.* [5] proposed a ultrasound physics-based simulation of the corresponding US image from the MR/CT volumes and used that to register with the live ultrasound images. Finally, in order to address the problem of small field of view of live ultrasound streams, Wein *et al.* [5] proposed using an extended field of view by sweeping the ICE catheter *in vivo* and reconstructing a larger field of view on-the-fly by constraining the catheter to lie on a linear trajectory model.

Contributions. In this paper, we develop an interventional guidance platform that includes a novel method to improve the accuracy and robustness of fusion between streaming live-3D ultrasound and pre-procedural MR images. An extended-FOV full-volume US image is used as an intermediary to register the live-3D ultrasound to the MR images. The large FOV of full-volume images and the similarity of contrasts and artifacts between full-volume US and live-3D ultrasound makes the proposed registration approach more robust, accurate, and efficient when compared to direct multimodal registration approaches. The interventional platform that integrates the proposed registration framework has been used and validated in three live animal experiments.

2 Methods and Materials

Clinical Workflow. Fig. 1 outlines the clinical workflow that was followed in using the proposed interventional guidance platform. The fusion of the imaging modalities is first bootstrapped using electromagnetic (EM) tracking system and then an image based registration is used to compensate for respiratory motion intra-operatively. First, prior to the procedure, the live-3D US image is pre-calibrated to the tracking

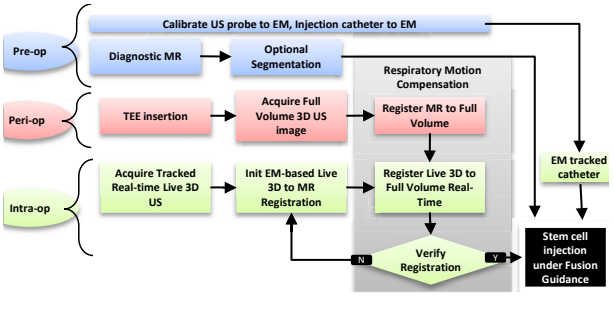


Fig. 1. Clinical Workflow for 3D US Guided Stem cell delivery

system by rigidly attaching 6-DOF EM sensors to the probe and calibrating using the method described in [6]. Second, the tip of the stem cell injection catheter is also pre-calibrated to the 5-DOF EM sensor mounted on it. During the procedure, pre-operative MR images are acquired with cardiac and respiratory gating. The specific MR volume that is used to segment the infarct is acquired at end-expiration and at a particular cardiac phase in diastole. In this work, we call that phase the 'infarct phase', which is determined by the MR imaging parameters. MR images are acquired with external fiducials stuck on the subject's body that are used for MR-EM registration. The injection targets are usually planned on the border zone between the infarct and healthy heart tissue.

During the procedure, a 6-degree of freedom (DOF) EM tracking sensor is rigidly mounted to the patient table and acts as a ground reference for the tracking system. MR images are registered to the EM frame of reference using fiducials. Following the MR-EM registration, TEE ultrasound probe is inserted into the esophagus. Live-3D ultrasound images are streamed live into the fusion workstation, wherein the fusion between MR and L3DUS is initialized using EM tracking. This initial registration is valid only at end-expiration and the 'infarct phase'.

The stem cell injection procedure, however, is performed in free breathing mode, during which the motion compensation algorithm is used to correct the registration for respiratory motion at every 'infarct' phase. In order to perform the motion compensation, an one time acquisition of full-volume 3D ultrasound volume is acquired using ECG gating at end-expiration. The MR images and the full-volume 3D US are registered using visual assessment. During the stem cell injection procedure, the live-3D US images are registered to the MR images using the full-volume 3D US using the algorithms described below. In addition, the EM tracked position of the injection catheter is displayed live within the motion compensated MR volume. Any gross motion of the US probe is also tracked, thereby providing continuous and robust tracking of the live-3D US volume in a motion compensated MR volume. The quality of the registration is monitored visually by the interventional cardiologist, and in the event of drift in registration, the registration is re-initialized using the EM based framework.

Real-time Respiratory Motion Compensation. The full-volume ultrasound (FVUS) is used as an intermediary in the registration between live-3D ultrasound (L3DUS) and MRI. Mathematically, ${}^{\mathrm{MR}}\mathbf{T}_{\mathrm{L3DUS}} = {}^{\mathrm{MR}}\mathbf{T}_{\mathrm{FVUS}} \bullet {}^{\mathrm{FVUS}}\mathbf{T}_{\mathrm{L3DUS}}$. The acquisition of full-volume ultrasound volumes is a feature that already exists in the Philips xMatrix$^{\mathrm{TM}}$ probes (Philips Healthcare, Bothell, WA). Full-volume imaging is an ECG gated acquisition of the 3D+Time volumetric acquisition of the entire heart in four heart beats. The 'infarct phase' in the full-volume image is identified using QLAB image analysis tool (Philips Healthcare, Bothell, WA) using the ECG gating values, and is manually registered to the MR image volume to yield a rigid transformation ${}^{\mathrm{MR}}\mathbf{T}_{\mathrm{FVUS}}$. The 'infarct phase' of the live 3D ultrasound stream is determined using a time-synchronized ECG gating signal that is output from the Ultrasound scanner via a dedicated A/D card (National Instruments, Austin, TX). Both the live-3D and the full-volume images are scan converted before registration.

Since the registration is performed only during the 'infarct phase' of the cardiac cycle, we use a rigid transformation to register the two ultrasound volumes. Mutual information (MI) is used as similarity metric in the registration [2]. A thresholding step

is performed on the ultrasound images prior to the mutual information computation to consider voxels only in areas with non-zero signal. The registration problem is framed as an optimization problem, $\arg\max_{\phi} \text{MI}\left(\texttt{L3DUS}, \texttt{FVUS} \circ \mathbf{T}(\phi)\right)$ where ϕ is a 6 parameter rigid transform vector parameterizing the rigid homogenous transformation matrix (${}^{\texttt{L3DUS}}\mathbf{T}_{\texttt{FVUS}}$) between full-volume ultrasound and live-3D ultrasound. The registration was implemented in an ITK framework [7]. A regular step size gradient descent optimizer was used for the optimization. The algorithm is integrated within the visualization platform with several user interfaces to control the optimizer behavior and the similarity metric computation, especially the number of histogram samples.

3 Experimental Design and Results

3.1 Phantom Validation

We conducted three phantom-based validation experiments to 1) validate the accuracy of EM-tracking based initialization, 2) measure the landmark registration accuracy, and 3) validate the motion tracking accuracy.

Accuracy of EM-Based Initialization. The accuracy of the EM based initialization was computed on a stationary heart phantom, which is a replica of a human heart, built in-house and constructed using poly-vinyl alcohol and doped to provide realistic visibility in both MR and ultrasound. The one time calibration of the live 3D US image coordinate to the 6-DOF EM sensor was performed with a calibration accuracy of 1.94 mm [6]. MRI images of the phantom were acquired using Philips Panorama 1T system with fiducials mounted on the phantom. The fiducials were used to register the MR to the EM frame of reference by localizing them in both coordinate systems. This approach resulted in a fiducial registration error (FRE) of 1.29 mm and target registration error (TRE) of 1.77 mm. It was also observed that keeping the fiducials too far apart decreased the accuracy due to EM distortions. Further, in order to measure US-MR fusion accuracy, seven home made thread-like fiducials were attached to the surface of the heart phantom. These were visible on both MRI and on 3D US images. By manually segmenting them in the 3D US and MRI images, and comparing the transformed points to the ones segmented in MRI, we estimated the US-MRI registration accuracy with a mean landmark accuracy 3.3 ± 0.22 mm.

Registration Accuracy Using Respiratory Motion Phantom. For the next two phantom validation studies, we designed a servomechanisms to move the heart phantom on a ramp to simulate respiratory motion (see Fig. 2). In addition, a new 'apex' was designed as a detachable slab that can fit custom made target samples to measure landmark accuracy. These samples are 4.5 mm in radius and are doped with magnevist and graphite to achieve realistic contrast in both MR and in ultrasound.

Image acquisition: An IE33 ultrasound scanner with X7-2T TEE probe (Philips Healthcare, Bothell, WA) was used to acquire both the full-volume and streaming ultrasound data sets. The pre-procedure MR was acquired on a GE 1.5T MR scanner (GE Healthcare, Waukesha, WI).

Experiments: The respiratory motion ramp had both translation and rotation stages. Ground truth motion was tracked using a dedicated 6-DOF EM sensor attached to the moving stage. The phantom was moved to three positions on the ramp at approximately 5, 10, and 15 mm from the baseline position; displacements that are representative of the amount of respiratory motion that we expect in a clinical setting [8]. At each position, three trials of the proposed registration was performed with the baseline registration being loaded before every trial. The full-volume based registration was performed using 1.6% of the number of pixels in the live-3D ultrasound image for histogram computation. The pixels were randomly sampled at each iteration of the registration. The amount of motion computed using the full-volume based registration was compared against the ground truth from the reference EM sensor. On an average from all trials and all positions, the translation error was 2.25 $\pm$ 1.96 mm and rotation error was 4.5 $\pm$ 1.98 degrees.

Landmark Accuracy Using Respiratory Motion Phantom. Another validation study was performed with the same phantom experimental set-up as described in the previous study. In this study, eight custom made fiducials were segmented from the MR images and the center of these segmentations were marked as targets. At each displacement of the phantom, three trials of full-volume US based registration was performed and the fusion between MR and live 3D was visualized. The centers of the fiducials were manually segmented from the US volumes (with MR visualization switched off to avoid bias) and transformed to the MR coordinate frame using the computed registration. These co-ordinates were then compared to the fiducials segmented in the MR coordinate frame. Since the FOV of the live-3D volumes is limited, only 3 fiducials were visible for computing landmark error. The average landmark registration error was 2.8$\pm$1.45 mm.

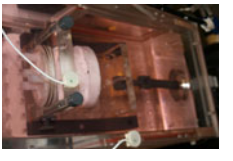


Fig. 2. Custom built heart phantom with multi-modality visibility employed for in vitro experiments, and screen shot of MR-US fusion

3.2 Animal Data Validation

We are currently evaluating the accuracy and robustness of the proposed fusion system in series of animal studies. In this paper, we present data from three anesthetized pigs with acute myocardial infarction that was created 2-4 days prior to the procedure. All studies were performed within the guidelines set by the Committee of Animal Research and Ethics at the University of Wisconsin Hospital, Madison, WI.

Image Acquisition: The baseline MR images were collected using the delayed hyper-enhancement MR protocol at breath-hold. The MR volumes are 256x256x20 with a resolution of 1.3x1.3x5 mm. Following the MR imaging, the pig is brought to the cath-lab where the TEE probe is inserted and full volume US images are acquired at breath-hold and end-expiration. After scan conversion, the size of the full-volume images are 224x208x208 with a pixel resolution of 0.5x0.5x0.48 mm. The live 3D images are streamed in real-time to the workstation at$\approx$ 20 Hz with a size of 112x48x112 with a pixel spacing of 0.7827x0.9403x0.9229 mm.

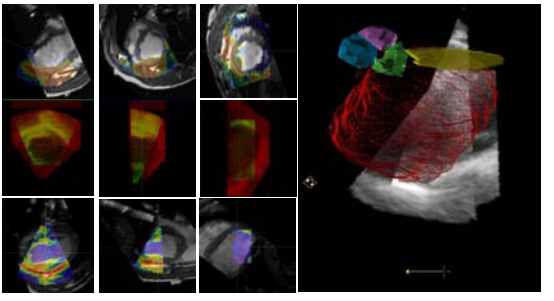


Fig. 3. Left:Top Row shows fusion of full-volume US and MR images, middle row shows results of intra-operative registration of live-3D US and full-volume US, bottom row shows fusion between live-3D and MR that is shown to the user. Right: 3D visualization of 2 multi-planar reformats (MPR) of the Live-3D ultrasound volume with the segmented MR images displayed as a mesh with aortic (yellow) and mitral valve leaflets also segmented to provide 3D context. The catheter is displayed as a yellow arrow seen in the bottom

The minimum number of histogram samples needed to achieve as stable registration varied between 1250 (0.06% of US image size) samples to 10000 samples (1.6% of US image size) depending on the amount of high-frequency details on the ultrasound images. The histogram parameters could be changed dynamically by the user during the procedure depending on the robustness of the registration. The time for registration varied between 350 milliseconds for 1250 to 3 seconds for 10000 samples on a 2.66GHz Dual Quad-Core processor with 4GB DDR memory. Example of a qualitative evaluation of the registration can be seen in Fig. 3, which shows good overlap between live-3D US and MR images in 2D and 3D visualizations.

***In vivo* Motion Compensation Accuracy.** One live-3D ultrasound image corresponding to the 'infarct phase' was selected manually from the live-3D US stream and analyzed off line. The ground truth registration between this live-3D image and the full-volume US image was selected manually with expert physician's guidance. Five hundred registration trials were performed, with initializations randomly misaligned around the ground truth registration with uniform distribution. For the capture range experiments, the translation and rotations were simultaneously varied by 15 mm and 15 degrees respectively with a uniform distribution. Similarly, for the accuracy related experiments, the translation and rotation shifts were simultaneously varied by 9 mm and 9 degrees respectively. The capture range was defined as the error below which >90% of the misalignments were able converge to an error < 3 mm. The number of histogram samples was set at 1.6% of image size of live-3D US. From data analyzed from 3 pig data sets, the average translation error compared to ground truth was 1.5 ± 1.7 mm, and rotation error was 4.9 ± 3.9 degrees, with overall registration accuracy of 2.2 ± 1.8 mm. These average values of accuracy and robustness to randomly varying initializations is indicative of the smoothness of the objective function between full-volume US and live-3D US. The average capture range for the algorithm was 15.3 ± 2.1 mm, which is greater than respiratory motion in humans [8].

***In vivo* Validation of Respiratory Motion Tracking.** Managing synchronization [9] and data frame rates of streaming data – ECG, ultrasound and EM — is a challenge

in interventional systems such as the one proposed in this paper. In order to test the proposed motion compensation system on reduced data rates, we tested its robustness to sub-sampling of US image streams in spatial and temporal dimensions. We established ground truth data by using a cardiac-gated 15 second long sequence high-resolution live-3D images acquired at 30 Hz (444 images in sequence) on the IE33 on free-breathing pig. The breathing rate was set at 17 breaths/min on the ventilator and the cardiac rate was 81 beats/min. Each frame was rigidly registered to the full volume 'infarct phase' in a sequential manner with the registration results from image n serving as an initialization to registration to image $n + 1$. Although the deformation between the full volume and live-3D is indeed non-rigid, a rigid registration transform is used to approximate the translation and rotation offset between the two images. The first initialization was done manually.

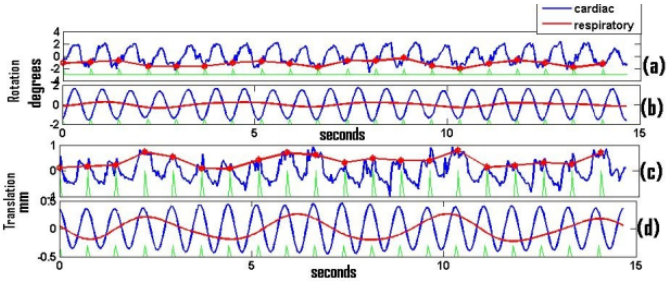


Fig. 4. Results of the motion compensation algorithm. Rotation and translation tracking of the cardiorespiratory motion of the heart respectively can be seen in (a) and (c) respectively. (b) and (d) show the underlying sinusoidal motion components of cardiac (blue) and respiratory (red) motion components

Data shown in Figs. 4(a) and (c) represents the computed rotation and translation motion of the heart that is used as the ground truth for this study. Figs. 4(b) and (d) show the respective sinusoidal components extracted using Fourier analysis of the motion tracking data. The blue curves are the measured rotation and translation values shown with mean subtracted for clarity of visualization. The red curves in these plots are samples of the tracking values at R-wave ECG trigger; and hence should ideally represent only respiratory component of cardiac motion. Note that the number of cardiac and respiratory cycles correlate with the cardiac and respiratory rate of the animal.

Table 1. Robustness to spatial and temporal decimation of live 3D streams

Temporal sub-sampling

Freq(Hz)	Rot err(deg)	Trans err(mm)
15	0.23	0.12
10	0.23	0.08
6	0.20	0.08
3	0.34	0.15

Spatial sub-sampling

Pix size red(%)	Rot err(deg)	Trans err(mm)
20	1.3	0.37
60	1.73	0.40
100	0.69	0.37
200	1.15	0.52

The stream of 444 images was sub-sampled in both spatial and temporal directions and the sequential registrations were performed on these sequences. The result was compared to the ground truth. Results in Table 1 show that the motion compensation is very robust to temporal

decimation of the streams with errors $<$ 1mm even for 3 Hz temporal acquisitions. In addition, the reduction of spatial resolution by 200% yields an error on only about 1.15 degrees and 0.52 mm suggesting that the algorithm is robust to sub-sampling in both spatial and temporal resolutions.

4 Discussion and Conclusion

A clinically translatable multi-modality fusion system for stem cell therapy has been proposed. The system integrates a novel motion compensation scheme that uses full-volume ultrasound imaging to enhance the accuracy, robustness, and speed of fusion between MRI and live-3D ultrasound. Validation on a moving heart phantom and preliminary data from animal studies have demonstrated the clinical feasibility of the proposed approach. Further *in vivo* validation of respiratory motion tracking and management of latencies in the system will be crucial to translate this technology to human clinical studies. Prospective registration schemes for respiratory motion compensation is another research direction to make this technology work in real-time.

References

1. Wollert, K., Drexler, H.: Does cell therapy work for myocardial infarction and heart failure work? Dialogues in Cardiovascular Medicine 14(1), 37–43 (2009)
2. Huang, X., Ren, J., Guiraudon, G., Boughner, D., Peters, T.M.: Rapid dynamic image registration of the beating heart for diagnosis and surgical navigation. IEEE Transactions on Medical Imaging 28(11), 1802–1814 (2009)
3. Sun, Y., Kadoury, S., Li, Y., John, M., Resnick, J., Plambeck, G., Liao, R., Sauer, F., Xu, C.: Image guidance of intracardiac ultrasound with fusion of pre-operative images. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part I. LNCS, vol. 4791, pp. 60–67. Springer, Heidelberg (2007)
4. King, A.P., Ma, Y.-L., Yao, C., Jansen, C., Razavi, R., Rhode, K.S., Penney, G.P.: Image-to-physical registration for image-guided interventions using 3-D ultrasound and an ultrasound imaging model. In: Prince, J.L., Pham, D.L., Myers, K.J. (eds.) IPMI 2009. LNCS, vol. 5636, pp. 188–201. Springer, Heidelberg (2009)
5. Wein, W., Camus, E., John, M., Diallo, M., Duong, C., Al-Ahmad, A., Fahrig, R., Khamene, A., Xu, C.: Towards Guidance of Electrophysiological Procedures with Real-Time 3D Intracardiac Echocardiography Fusion to C-arm CT. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5761, pp. 9–16. Springer, Heidelberg (2009)
6. Lang, A., Parthasarathy, V., Jain, A.: Calibration of 3D Ultrasound to an Electromagnetic tracking system. In: D'hooge, J., Doyley, M.M. (eds.) Proceedings of the SPIE Medical Imaging 2011: Image Guided Procedures. SPIE, vol. 7968 (2011)
7. Ibanez, L., Schroeder, W., Ng, L., Cates, J.: The ITK Software Guide, 2nd edn (2005)
8. Shechter, G., Ozturk, C., Resar, J.R., McVeigh, E.R.: Respiratory motion of the heart from free breathing coronary angiograms. IEEE Trans. Med. Im. 23(8), 1046–1056 (2004)
9. Sundar, H., Khamene, A., Yatziv, L., Xu, C.: Automatic image-based cardiac and respiratory cycle synchronization and gating of image sequences. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5762, pp. 381–388. Springer, Heidelberg (2009)

Towards Intra-operative Prostate Brachytherapy Dosimetry Based on Partial Seed Localization in Ultrasound and Registration to C-arm Fluoroscopy

Mehdi Moradi[1,⋆], S. Sara Mahdavi[1], Sanchit Deshmukh[2], Julio Lobo[1], Ehsan Dehghan[3], Gabor Fichtinger[3], William J. Morris[4], and Septimiu E. Salcudean[1,⋆]

[1] Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, Canada
{moradi,saram,juliol,tims}@ece.ubc.ca
[2] Indian Institute of Technology, Bombay, India
[3] School of Computing, Queen's University, Kingston, ON, Canada
[4] British Columbia Cancer Agency

Abstract. Intraoperative dosimetry during prostate brachytherapy is a long standing clinical problem. We propose a novel framework to address this problem by reliable detection of a subset of seeds from 3D transrectal ultrasound and registration to fluoroscopy. Seed detection in ultrasound is achieved through template matching in the RF ultrasound domain followed by thresholding and spatial filtering based on the fixed distance between stranded seeds. This subset of seeds is registered to the complete reconstruction of the implant in C-arm fluoroscopy. We report results, validated with a leave-one-needle-out approach, both in a phantom (average post-registration seed distance of 2.5 mm) and in three clinical patient datasets (average error: 3.9 mm over 113 seeds).

1 Introduction

Low dose rate brachytherapy is a minimally invasive therapeutic procedure for prostate cancer that has rapidly gained acceptance due to highly successful clinical results. In this procedure, a number of small radioactive sources or seeds (^{125}I or ^{103}Pd) are permanently implanted into the prostate using brachytherapy needles. The aim is to deliver a sufficient radiation dose to kill cancerous tissue while limiting the dose in radio-sensitive regions such as the bladder, urethra and rectum. Transrectal ultrasound (TRUS) is used to intraoperatively guide the transperineal insertion of needles. As a result of prostate edema, motion of the gland due to needle forces, and possible intra-operative changes to the plan due to various factors such as interference with the pubic arch, the locations of the implanted seeds do not necessarily match with the initial treatment plan. Hence, for quality assurance, intra-operative dosimetry is highly beneficial.

⋆ Corresponding authors.

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 291–298, 2011.

Researchers have approached ultrasound-based seed detection [1]. However, accurate seed localization based on ultrasound has proven to be a very difficult task due to clutter from other highly reflecting objects such as calcifications resulting in false positive appearances, seed specularity and shadowing, and limited field of view. Even when hand-segmented, up to 25% of the seeds remain hidden in ultrasound images [2]. Therefore, C-arm fluoroscopy is commonly used for visual assessment of the implanted seeds. However, the prostate gland is not discernible in fluoroscopy images. Fusion of the fluoroscopy images and ultrasound is therefore considered as a possible solution [3–6]. If complete seed localization and implant reconstruction from fluoroscopy images is available, the registration of the result to ultrasound will enable dosimetry. In recent years, the fluoroscopy reconstruction problem has been extensively addressed. Given 3-5 fluoroscopy images, and the relative pose of the C-arm in each acquisition, the back-projection technique can be used to reconstruct the 3D implant [7]. For registration, attaching fiducial markers to the ultrasound probe [8], using the ultrasound probe itself as a fiducial [4], or using the seeds as fiducials [9] have been mentioned. Due to patient and equipment motion between the acquisition of ultrasound and fluoroscopy, registration based on static markers is not reliable. Furthermore, the use of fiducial markers is an unwelcome addition to the ordinary setup in the operating room due to time and space limitations.

In this work we propose using a subset of seeds extracted from ultrasound images to perform point-based registration between the seed clouds from fluoroscopy and 3D ultrasound. We present several technical innovations. Instead of conventional B-mode ultrasound, we use RF signals processed to enhance seed contrast. Template matching with a variety of *in vivo* and *ex vivo* seed templates is reported. To enable dosimetry, we have devised a two stage strategy consisting of first 2D registration of needle projections from the ultrasound and fluoroscopy, followed by the 3D registration of only the seeds in the matched needles. We provide the results of this approach on both clinical and phantom data. This novel methodology targets a complicated and long standing problem, with no addition to the routine therapeutic procedure. We show that our method has good promise to address this clinical challenge.

2 Methods

The outline of our methodology is presented in Figure 1. For reconstruction of the implant in fluoroscopy, we implemented and used the method described in [7]. The steps for acquiring the fluoroscopy reconstruction included C-arm pose estimation from rotation angle and compensation for sagging, followed by back-projection of the seeds, and finally seed to needle assignment using a minimum cost flow (MCF) algorithm. The outcome was validated in terms of number of seeds, needles, and seed to needle assignments based on the brachytherapy plans, both in patient and phantom datasets. In this article we focus on ultrasound-based partial seed detection, needle matching and registration from ultrasound to fluoroscopy.

2.1 Seed Detection

3D Ultrasound Setup and Data: We developed a 3D ultrasound system based on a brachytherapy stepper (EXII, CIVCO Medical Solutions) modified by motorizing the cradle rotation. The sagittal array of a dual plane linear/microconvex broadband $5-9$ MHz endorectal transducer (Ultrasonix) was used. RF data was recorded at a frame rate of 42 fps, during the probe rotation from -45° to 50° (0° corresponded to the probe aligned with the central sagittal cross section of the prostate gland). 2D frame size was 5×5.5 cm. We present the results of our work on data collected immediately after seed implantation in the OR, from three brachytherapy patients in Vancouver Cancer Center. We also present data from a CIRS Model 053 tissue-equivalent ultrasound prostate phantom (CIRS, Inc., Norfolk, VA). For this phantom, a plan consisting of 135 seeds and 26 needles was created which was carried out by a radiation oncologist.

RF Signal Processing: In order to improve the seed to background contrast, we averaged the signal power over windowed blocks of the RF signals. In other words, we replaced a segment of length n at depth d of an RF line with the reflected power P_d computed as:

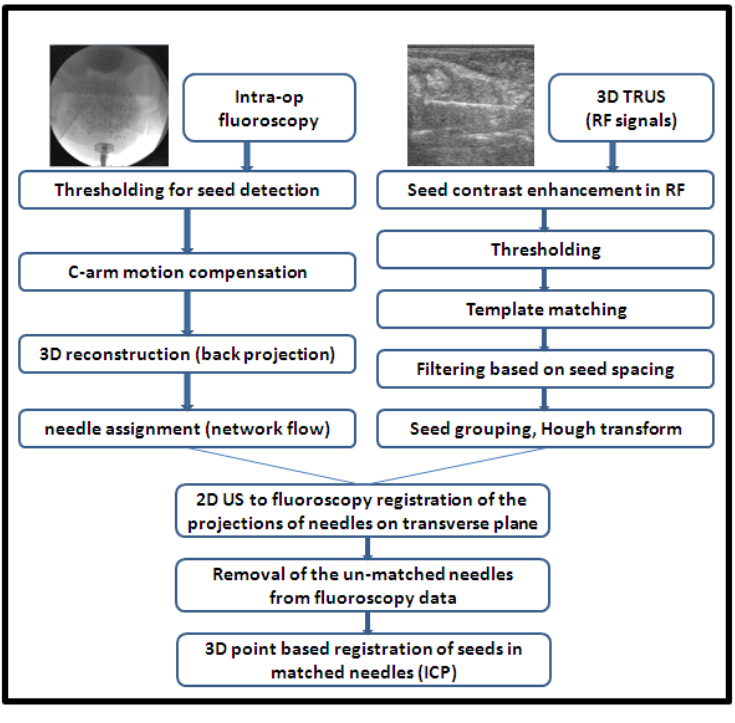


Fig. 1. Workflow of the seed detection method

$$P_d = \frac{\sum_{k=1}^{n} w(k)x(k)^2}{n} \quad (1)$$

where $x(k)$ $(k = 1, ..., n)$ are the samples in the RF segment, and $w(k)$ are Hamming window weights. This step was applied with n =10 and a 50% overlap between consecutive blocks. We have previously reported that this process doubles the contrast to noise ratio (CNR) between the seed regions and the background [10]. Furthermore, the resulting five fold reducion in the size of the RF signals improved the speed of the template matching step.

Seed Templates: Simple thresholding of the contrast enhanced RF data results in a large number of seed candidates and seed-like clutter. In order to reduce false positive detections, we computed the normalized cross correlation of the seed regions with seed templates. We experimented with three groups of templates and will provide a comparison.

Ex vivo templates: We created a 3D template acquired by imaging a seed in water, placed parallel to the probe at the center of the ultrasound probe. The 3D template was formed by rotating the probe and acquiring 21 sagittal images.

In vivo templates: In the clinical situation, the existence of background tissue, blood and edema significantly alters the appearance of seeds on ultrasound images compared to the described *ex vivo* templates, resulting in low normalized cross correlation values. Therefore, we created a second set of templates extracted from *in vivo* data. These were extracted from different cases to ensure that the template extracted from a specific patient dataset, was not used for seed detection in that dataset. They were acquired by manually clicking the center of a visually distinct seed in *in vivo* 3D data.

In vivo and *ex vivo* two-seed templates: We also created templates, both *in vivo* and *ex vivo*, in which the template area contained two seeds. The two-seed templates were examined as a potential solution to reduce the number of false positives, given the fact that in stranded brachytherapy performed in our institution, the distance between seeds is fixed at 1 cm with very few exceptions. Therefore, the existence of two strong seed candidates, 1 cm apart, is an indication of a true detection.

NCC-Based 3D Template Matching: The normalized cross correlation of the 3D template g and a cropped area of the image f equal in size with g and centered at location (i, j, k) can be computed as:

$$f \star g(i,j,k) = \frac{(\sum_{x,y,z} f_{i,j,k}(x,y,z) - \overline{f_{i,j,k}}).(g(x,y,z) - \overline{g})}{\sqrt{[\sum_{x,y,z}(f_{i,j,k}(x,y,z) - \overline{f_{i,j,k}})^2 \sum_{x,y,z}(g(x,y,x) - \overline{g})^2]}} \quad (2)$$

where x, y and z represent the directions in the image coordinate system. Computation of $f \star g$ results in a new image of the same size as f, with values in range of [0,1] with largest values representing the centers of areas most similar to the template. The frequency domain implementation of NCC was completed in under six minutes on a regular PC with MATLAB for template size of 30×60×21, and

image size of 128×258×391. Note that NCC is not invariant to scale. However, in our case, the images and the templates were acquired with similar imaging parameters, therefore the scales matched.

Thresholding, Spatial Filtering, and Grouping in the NCC Image: The NCC image was thresholded. Starting from the point with the highest NCC, a neighborhood of the size of the seed was cleared around each non-zero voxel. This was necessary because each seed consists of several bright voxels, while we need a single voxel to represent the seed. The remaining voxels were grouped into needles using the Hough transform [11] followed by eliminating single seeds that could not be grouped into lines. Using the knowledge of the fixed 1 cm distance between the seeds in our data, we applied an additional trimming step. On each needle, starting from the seed with the highest NCC value, any other seed candidate that was within 0.8 cm was removed.

2.2 Matching and Registration

After applying a transformation that matched the centers of mass of the two datasets, we applied a 2D needle matching process. This provided an initial alignment and reduced the risk of local minima due to the unbalanced number of seeds in ultrasound and fluoroscopy. Matching was performed by applying a 2D rigid registration between the needle projections on the transverse plane passing through the prostate mid-gland in the fluoroscopy implant. Assuming that X is the set of n projection points from ultrasound, and Y is the set of m projection points from fluoroscopy, the rotation and translation parameters of the transformation T were found to minimize $\sum_{i=1:n} d_c[T(X_i), Y]$ where $d_c[T(X_i), Y]$ is the Euclidean distance of the ultrasound projection point X_i from its closest match in the point set Y. After the matching step, the fluoroscopy needles without a match in ultrasound were removed and the standard 3D point-based Iterative Closest Point (ICP) registration algorithm [12] was applied.

3 Results and Conclusions

We quantified the outcome of our ICP seed registration method based on 1) the post-registration distances between ultrasound seeds and their closest fluoroscopy counterparts, 2) the stability of needle matches and the recorded registration errors subject to the removal of any of the detected needles.

Registration Errors: Table 1 presents the results of ultrasound seed detection and registration for the CIRS phantom. The best outcome was achieved when the two-seed *ex vivo* template set was used. The post-registration seed distances from ultrasound to the closest matching fluoroscopy seed was 2.48 mm. Note that the best registration result was obtained when the lowest number of seeds (73 out of 135, 17 needles out of 26) were detected. The result of the 2D matching and 3D registration of the seed clouds for the phantom data, using the two-seed *ex vivo* template, is depicted in Figure 2.

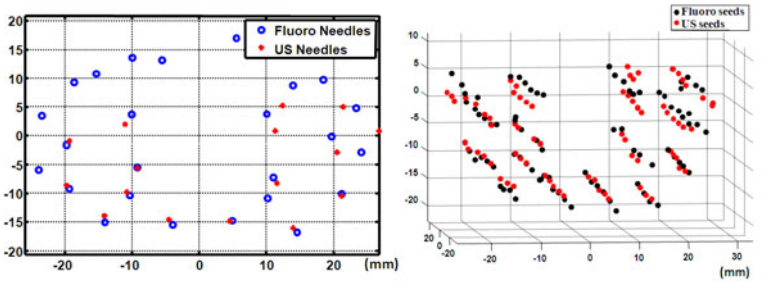


Fig. 2. Phantom results: 2D matching of needle projections in transverse plane (left), 3D seed cloud registration using the matched needles from fluoroscopy and ultrasound

Table 1. Phantom result: # of detected seeds, registration errors with all detected needles, and average registration error for the leave-one-needle-out validation test

Phantom	dist. er.(mm) all needles	dist. er. (mm) L.O.O	# fl seeds Total	# fl seeds, in matching needles	# of detected seeds in ultrasound
1-seed, *ex vivo* templ.	2.69±2.03	2.71±1.98	135	95	86
1-seed, *in vivo* templ.	4.33±2.21	4.40±2.24	135	101	95
2-seed, *ex vivo* templ.	2.48±1.52	2.48±1.41	135	82	73

Table 2 presents the results for the three patient datasets. The one-seed and the two-seed templates did not result in significantly different error values. We obtained errors of 3.36 mm, 3.73 mm, and 4.76 mm for the three cases on *in vivo* templates. With *ex vivo* templates, we witnessed a decrease in the number of detected seeds for all cases and an error increase in two patients with a slight improvement in case 3 in terms of registration error (to 4.38, 4.08 and 4.22 mm). For case 2, Figure 3 illustrates the results of 2D matching and 3D registration. In both Figures 2 and 3 one can see that the unmatched needles tend to be from the anterior side (top of the images), while the posterior seeds that are closest to the probe are accurately detected. This is likely due to signal attenuation and the depth dependent reduction in the resolution of our 3D ultrasound system.

Leave-one-needle-out Validation of the Registration Process: In order to study the stability of our matching and registration procedure, we also ran a leave-one-needle-out experiment. For each patient and the phantom case, assuming that n ultrasound needles were identified, we repeated the matching and registration procedure n times, each time with $n-1$ needles. This amounted to the removal of three to seven (10% to 20%) of the seeds in each round for patient cases. The idea is that if the registration is valid, and not just a local minimum, the removal of any specific needle should not drastically change the outcome.

The average of the errors in the leave one out experiments (column 3 in Tables 1 and 2) were close to the errors when all detected seeds were used in the registration step (column 2). We also examined the stability of the needle

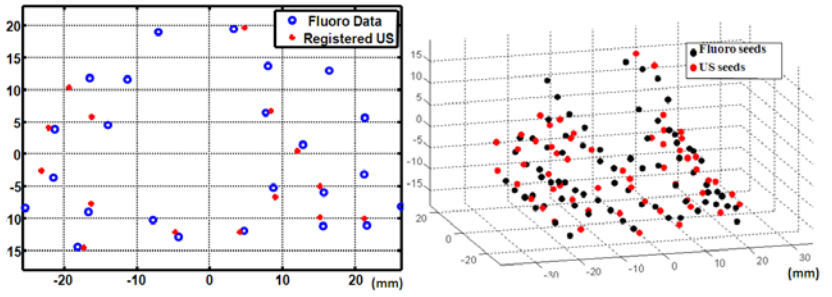


Fig. 3. A patient case: 2D matching (left), registered seeds in matched needles (right)

Table 2. Number of the detected seeds, registration errors and leave-one-needle-out errors for the 3 patient datasets

Case	dist. er.(mm) all needles	dist. er. (mm) L.O.O	# fl seeds Total	# fl seeds, in matching needles	# of detected seeds in ultrasound
P1 1-seed *in vivo* templ.	3.72±1.86	3.85±1.96	102	51	35
P1 2-seed *in vivo* templ.	3.36±1.78	3.44±1.95	102	57	37
P1 *ex vivo* templ.	4.07±2.10	4.38±1.87	102	30	16
P2 1-seed *in vivo* templ.	3.98±2.50	3.77±2.27	115	74	56
P2 2-seed *in vivo* templ.	3.73±1.86	3.79±1.94	115	71	49
P2 *ex vivo* templ.	3.74±2.8	4.08±2.58	115	55	37
P3 1-seed *in vivo* templ.	4.76±1.87	4.90±2.12	100	35	27
P3 2-seed *in vivo* templ.	4.88±1.95	4.79±2.25	100	33	23
P3 *ex vivo* templ.	4.52±1.65	4.22±1.49	100	26	16

matching step subject to removal of needles. In the case of the phantom data, and case 1 among the patient datasets, it was noted that regardless of which needle was removed in the leave one out experiment, the matched fluoroscopy needle for the rest of ultrasound needles remained the same (number of detected needles: n=17). In patient cases 2 and 3, on average, the match for one needle changed due to the removal of a needle (n =14 and n =12).

4 Conclusions

We showed that it is feasible to use contrast enhanced RF ultrasound data, template matching, and spatial filtering to detect a reliable subset of brachytherapy seeds from ultrasound to enable registration to fluoroscopy. Real-time implementation requires the matching process to be computationally improved though parallelization to enhance the current computation time of around six minutes. More robustness analysis and additional registration approaches will be implemented in future work. The 3D ultrasound system used in this work acquired sagittal images while rotating radially. This results in a significant decrease in

spatial resolution at increasing distances from the probe. The use of a comprehensive depth dependent set of templates can improve our results. It is also possible to perform the ultrasound seed detection after only a part of the implant, for example the top row of needles, is completed. This will reduce the shadowing effects. Clinical data is being acquired to test these possibilities.

Acknowledgments. Dr T. Pickles and Mr N. Chng (data acquisition); Dr X. Wen (RP imaging codes). M. Moradi held an NSERC PDF and the US DoD PCRP Award W81XWH-10-1-0201, S. Mahdavi was supported by the Prostate Cancer Foundation of BC, G. Fichtinger was supported as Cancer Care Ontario Research Chair, S. E. Salcudean was supported by NSERC and CIHR.

References

1. Mamou, J., Ramachandran, S., Feleppa, E.J.: Angle-dependent ultrasonic detection and imaging of brachytherapy seeds using singular spectrum analysis. Journal of the Acoustical Society of America 123(4), 2148–2159 (2008)
2. Jain, A.K., Zhou, Y., Mustufa, T., Burdette, E.C., Chirikjian, G.S., Fichtinger, G.: Matching and reconstruction of brachytherapy seeds using the hungarian algorithm (MARSHAL). In: SPIE Medical Imaging. SPIE, vol. 5744, pp. 810–821 (2005)
3. Fallavollita, P., Karim-Aghaloo, Z., Burdette, C.E., Song, D.Y., Abolmaesumi, P., Fichtinger, G.: Registration between ultrasound and fluoroscopy or CT in prostate brachytherapy. Med. Phys. 37(6), 2749–2760 (2010)
4. French, D., Morris, J., Keyes, M., Salcudean, S.E.: Real-time dosimetry for prostate brachytherapy using TRUS and fluoroscopy. In: Barillot, C., Haynor, D.R., Hellier, P. (eds.) MICCAI 2004. LNCS, vol. 3217, pp. 983–991. Springer, Heidelberg (2004)
5. Su, Y., Davis, B.J., Furutani, K.M., Herman, M.G., Robb, R.A.: Seed localization and trus-fluoroscopy fusion for intraoperative prostate brachytherapy dosimetry. Comput. Aided Surg. 12(1), 25–34 (2007)
6. Todor, D.A., Zaider, M., Cohen, G.N., Worman, M.F., Zelefsky, M.J.: Intraoperative dynamic dosimetry for prostate implants. Physics in Medicine and Biology 48, 1153–1171 (2003)
7. Dehghan, E., Lee, J., Moradi, M., Wen, X., Fichtinger, G., Salcudean, S.E.: Prostate brachytherapy seed reconstruction using C-arm rotation measurement and motion compensation. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6361, pp. 283–290. Springer, Heidelberg (2010)
8. Zhang, M., Zaider, M., Worman, M., Cohen, G.: On the question of 3D seed reconstruction in prostate brachytherapy: the determination of x-ray source and film locations. Phys. Med. Biol. 49(19), N335–N345 (2004)
9. Tutar, I.B., Gong, L., Narayanan, S., Pathak, S.D., Cho, P.S., Wallner, K., Kim, Y.: Seed-based transrectal ultrasound-fluoroscopy registration method for intraoperative dosimetry analysis of prostate brachytherapy. Med. Phys. 35(3), 840–848 (2008)
10. Wen, X., Salcudean, S.E., Lawrence, P.D.: Detection of brachytherapy seeds using 3-D transrectal ultrasound. IEEE Trans. Biomed. Eng. 57(10), 2467–2477 (2010)
11. Duda, R.O., Hart, P.E.: Use of the Hough transformation to detect lines and curves in pictures. Communications of the ACM 15, 11–15 (1972)
12. Besl, P.J., McKay, N.D.: A method for registration of 3-D shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence 14(2), 239–256 (1992)

Effects of Ultrasound Section-Thickness on Brachytherapy Needle Tip Localization Error

Mohammad Peikari, Thomas Kuiran Chen, Andras Lasso,
Tamas Heffter, and Gabor Fichtinger

School of Computing, Queen's University, Kingston, Ontario, K7L 3N6, Canada
{mpeikari,chent,lasso,heffter,gabor}@cs.queensu.ca

Abstract. Purpose: Ultrasound section-thickness is the out-of-plane beamwidth causing major roles in creating image artifacts normally appearing around the anechoic areas. These artifacts can introduce errors in localizing the needle tips during any ultrasound-guided procedure. To study how section-thickness and imaging parameters can affect observing and localizing needle tips, we have conducted a typical calibration setup experiment. Method: Multiple needles were inserted orthogonal to the axial image plane, at various distances from the transducer. The experiment was conducted on a brachytherapy stepper for a curvilinear transrectal-ultrasound probe. Result: Experiments demonstrated that the imaging parameters have direct impacts on observing needle tips at different axial locations. They suggest specific settings to minimize the imaging artifacts. Conclusion: The ultrasound section-thickness and side lobes could result in misjudgment of needle insertion depth in an ultrasound-guided procedure. A beam profile could assist in considering the likelihood of position errors, when the effects of side lobes are minimized.

1 Introduction

Ultrasound (US) imaging is ubiquitous in intra-operative surgical guidance. It has been discussed in the literature [1–5] that the US images may contain certain artifacts caused by the section-thickness (elevation beamwidth) of the beam, orthogonal to both the axial and lateral beam axes. These artifacts may conceal tissue structures and may lead to incorrect medical diagnosis [1]. The motivating application for the present work is transrectal ultrasound (TRUS) guided prostate cancer brachytherapy, a procedure that entails the permanent implantation of small radioactive capsules through hollow needles into the patient's prostate in order to eradicate the cancer with radiation. During the brachytherapy procedure, the physician uses ultrasound slices to visualize the current position of the needle tip that appears as a bright spot in the image. Section-thickness is a unique problem for TRUS-guided brachytherapy because the needles are perpendicular to the ultrasound image slice. The nonuniform section-thickness causes error in localizing the needle tip and thus lead to inaccurate needle placement and ultimately to suboptimal deposition of the radioactive dose. The

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 299–306, 2011.

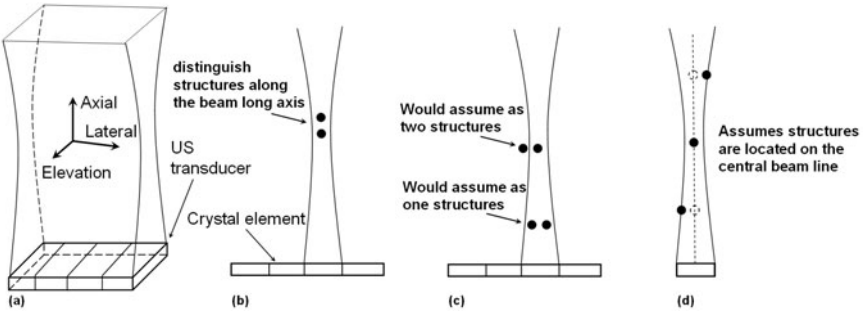


Fig. 1. (a) US beam pattern in *Axial, Lateral,* and *Elevation* axes. (b) Axial resolution, (c) Lateral resolution, and (d) Elevation resolution

objective of this paper is to quantify the needle tip placement error in during the brachytherapy procedure.

The quality of an US image primarily depends on three factors. Axial and lateral resolutions, which is the ability of the US device to distinguish between two structures along the axial and lateral direction, respectively. The third factor is the so called elevation resolution, where the US device assumes that all received echoes originate from structures situated precisely on the central line of the US beam [6], as depicted in Fig. 1 and in more details in Fig. 2. Figure 2(a) and (b) show the cross-section of an unfocused beam pattern and its corresponding A-mode echoes generated by reflecting materials in the medium. The first reflected beams, A-C, correspond to the three objects located at the same depth (D) from the transducer within the near-field of the US beam. Since the strength of the US beam is at its peak at the center and it decreases gradually toward the side edges of the beam [1], the maximum echo amplitude (A) corresponds to point A on the central beam line and the minimum echo corresponds to point C near the side edges of the beam. The same concept applies to reflecting objects D-F, except that the overall US beam intensity decreases farther away from the transducer. Since echoes from the same axial distance and lateral position of the US beam are received by the transducer at the same time, they are absorbed simultaneously. All echoes received at the same time are summed and interpreted as a single object located on the US central beam line [1]. As a consequence, echoes from an object located along the side edge of the US beam appear to be originating from a non-existing virtual reflector on the central region of the US beam, shown as point G in Fig. 2(b). Putting it simply, reflectors along the side edge of the beam do not appear at their true position.

The US section-thickness has been measured using phantoms incorporating inclined surfaces and multiple filaments [1–4]. Recently, we have constructed a device to measure the section-thickness of side firing TRUS probes [5]. We used the same replica of this device for measuring needle localization and placement

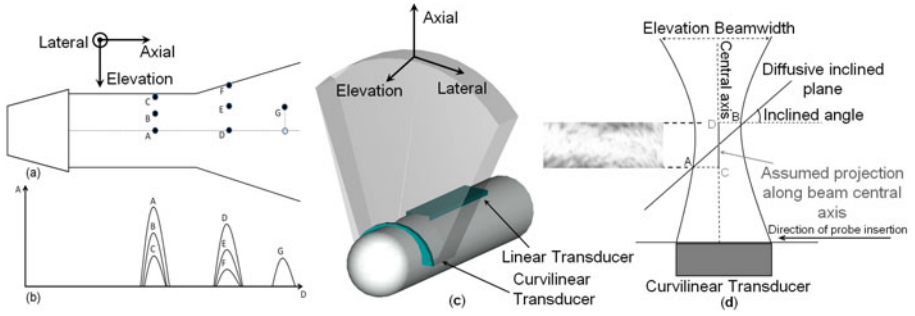


Fig. 2. (a)-(b) Beam pattern of an unfocused B-mode and A-mode US linear array with corresponding point reflectors [1]. (c) Axial, lateral and elevational axes convention with respect to the TRUS beam pattern. (d) Section-thickness estimation principle.

errors caused by the section-thickness of the TRUS beam in a standard clinical brachytherapy setup. To the best of our knowledge there has been no published work on this subject in the open literature.

2 Methodology

2.1 Experimental Setup

To generate a beam profile for the TRUS probe at 6MHz central frequency, we moved the probe back and forth (along the elevation axis) to acquire US images of the inclined-plane device (Fig. 3(a)). This is performed when the TRUS probe transducer and profiling phantom were inserted into a clear water bath.

To observe needle tip effects (appeared as bright spots in Fig. 4(c)) on the US images, 6 needles were inserted through holes of a brachytherapy stepper grid template as shown in Fig. 3(b). The needles were placed at the grid's central holes to ensure the distorted parts of the image (along the sides) does not have any influence in our measurements. For each needle, we moved the TRUS probe back and forth (along the elevation axis) in a water bath until the reflection from the needle tips appear on the US image as they intersect with the US beam's boundary. The corresponding probe depth were then recorded for further analysis.

2.2 US Beam Profile Estimation

To have an estimate of localization error for a TRUS probe, the US main beam thickness for all depths of the US beam (effective imaging region) from the transducer is found. We do this for the curvilinear transducer by using the same principle first explained in [1]. According to this approach (Fig. 2(d)), as the US beam propagates through the medium it first intersects with point A (nearest point to the transducer) on the inclined diffusive surface (45° to the beam). Similarly, the last point with which the US beam intersects would be the point

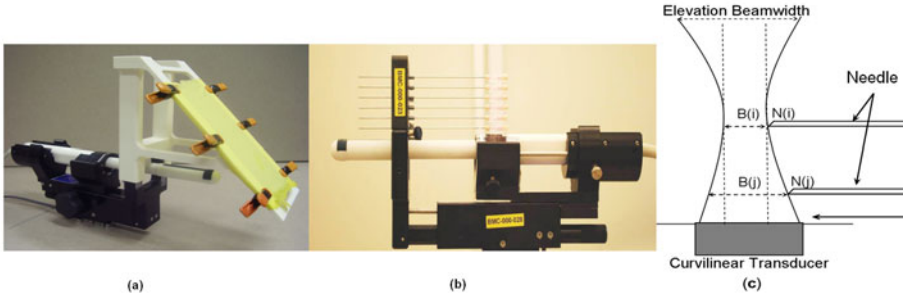


Fig. 3. (a) The beam profiling phantom mounted on a Burdette Medical Systems stepper, (b) Needle insertion setup, (c) Needle offset measurement principle

B (farthest point to the transducer). The sound echoes return to the transducer, however, since the US device assumes all the echoes received to the transducer along the elevation axis of the beam are from the structures on the beam's central line, the line AB would be assumed as line CD to the device. Hence the TRUS image would include a thick bright band which the thickness approximately represents the US main beam thickness. A set of TRUS images with different imaging parameters were collected at 6MHz for all depths of the imaging region. The artifacts (bright bands) were segmented manually from the images and their distances to the US transducer (position of the band) were taken as the depth measurements. A beam profile of US main beam thickness versus axial distance from the transducer is then plotted as shown in Fig. 4(b) and Fig. 5(a),(c), and (e). Figure 4(a) shows a subset of these measurements.

2.3 Needle Offset Measurement on US Images

To calculate the needle insertion offsets for every experiment, we set the closest needle to the transducer (needle No. 1) as the reference needle and subtract the depths of the other observed needle positions from that of the reference needle, as shown in Fig. 3(c). The subtracted values represent the amount of divergence or convergence of the beam pattern with respect to the reference needle. Hence the section-thickness relationship between every two inserted needles is defined as:

$$B(j) = |(N(j) - N(i))| \times 2 + B(i)$$

Where i and j are the two axial depths where the needles are inserted; $N(i)$ and $N(j)$ are the needle insertion depths at axial depths i and j respectively; and $B(i)$ and $B(j)$ are the US main beam thickness at the corresponding i and j axial depths respectively.

3 Results and Discussion

In order to observe the effects of the US device imaging parameters on the beam pattern we performed a series of needle insertion tests for 27 combinations of US

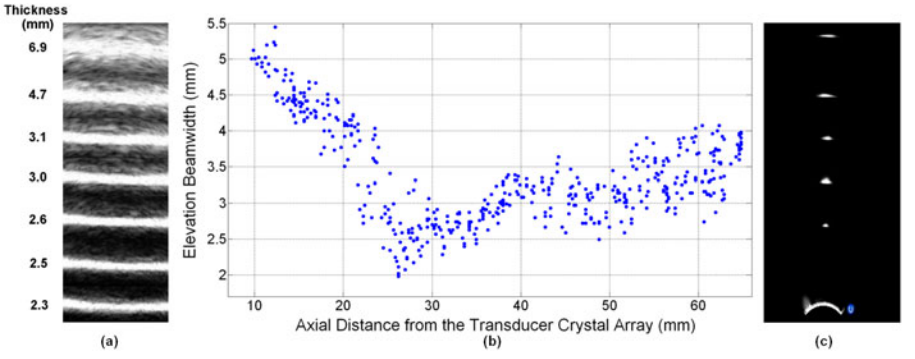


Fig. 4. (a) A subset of elevation main beam thickness measurements from 28mm to 90mm axial distance, at gain= 55%, dynamic range= 78dB, and power= 0. (b) TRUS elevation beam profile at 6MHz central operating frequency from 10mm to 65mm axial distance, at gain= 55%, dynamic range= 78dB, and power= 0. (c) Needle tip appearance in a TRUS image at gain= 0%, dynamic range= 15dB, and power= −7.

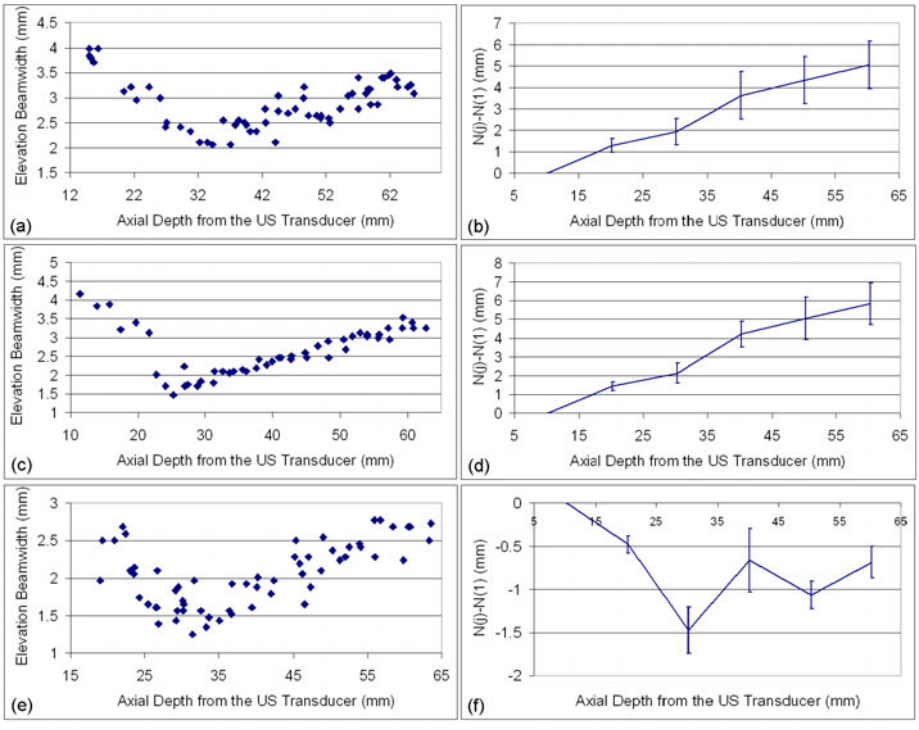


Fig. 5. (a) US beam profile for gain= 100%, dynamic range= 50dB, and power= 0. (b) Average and std. of $N(j) - N(1)$ for gain= 100%. (c) US beam profile for gain= 50%, dynamic range= 50dB, and power= 0. (d) Average and std. of $N(j) - N(1)$ for gain= 50%. (e) US beam profile for gain= 0%, dynamic range= 50dB, and power= 0. (f) Average and std. of $N(j) - N(1)$ for gain= 0%.

Table 1. Summary of the needle insertion offsets for different US settings

Gain (%)	Dyn. Range (dB)	Power	N(j)-N(1) (mm)					
			j=1	j=2	j=3	j=4	j=5	j=6
0	15	0	0	-0.354	-1.134	0	-0.992	-0.567
		-4	0	-0.354	-1.204	-0.07	-1.134	-0.637
		-7	0	-0.496	-1.13	-0.921	-1.063	-0.708
	50	0	0	-0.496	-1.559	-0.637	-1.275	-0.779
		-4	0	-0.567	-1.417	-0.85	-1.134	-0.637
		-7	0	-0.496	-1.559	-0.637	-1.275	-0.779
	100	0	0	-0.425	-1.488	-0.992	-1.134	-1.063
		-4	0	-0.567	-1.701	-0.851	-1.134	-0.567
		-7	0	-0.637	-1.77	-0.708	-0.708	-0.425
50	15	0	0	1.913	3.047	5.174	6.449	7.796
		-4	0	1.488	1.913	4.748	5.17	6.024
		-7	0	1.204	2.26	4	4.181	6.024
	50	0	0	1.771	2.835	5.315	6.024	7.016
		-4	0	1.913	2.551	4.536	5.528	6.591
		-7	0	0.992	2.197	1.913	2.976	4.181
	100	0	0	1.488	2.338	4.6	5.244	6.237
		-4	0	1.488	2.409	4.394	4.465	6.166
		-7	0	1.204	1.771	3.827	6.449	5.741
100	15	0	0	1.134	2.055	4.394	5.244	5.457
		-4	0	1.134	2.126	3.402	4.252	4.748
		-7	0	1.275	1.7	2.551	2.976	4.394
	50	0	0	1.488	2.409	4.6	6.095	5.67
		-4	0	1.204	1.772	3.118	3.402	5.174
		-7	0	1.417	1.134	3.331	3.969	3.543
	100	0	0	1.13	1.488	3.685	4.394	5.032
		-4	0	1.204	1.559	4.252	4.252	3.969
		-7	0	1.063	0.779	2.551	3.402	4.04

gain (0, 50, and 100%), dynamic range (15, 50, and 100 dB) and power (0, -4, and -7). The results are shown in Fig. 5 and Table 1.

According to Fig. 5(b)-(d), when gain=50% or gain=100%, the needle offsets are monotonously increasing, which indicates the beam diverges constantly. However, this is not the case when comparing with the beam profile generated at the same settings (Fig. 5(a)-(c)). The beam profile pattern shows that the US beam converges up to a focal point and diverges right after that. This contradiction in the beam profile and needle insertion offset plots could be because of the US side lobes artifacts.

Side lobes consist of multiple low-intensity off-axis ultrasound beams that produce image artifacts due to the error in localizing the returning echoes within the main US beam [6, 7]. If a highly reflective structure is encountered, it will be wrongly positioned in the image along the main US beam [6]. When the gain is set to high, the energy assigned to the US side lobes increases and hence their effects on the TRUS images increase. Hence, during needle insertion, the needle tips are first intersect with the side lobe energies and their echo artifacts are shown as if they are intersected with the main US beam. This clearly shows that the US main lobe thickness and the side lobe artifacts together might have large effects on localizing needle tips and objects within the TRUS images.

On the other hand, when gain=0%, the side lobe energies are set to be minimum and the needle offsets are less than zero (Fig. 5(f)) which indicates that the

beam constantly converges up to a focal zone (around 30 mm from the transducer). The beam pattern starts diverging quickly after the focal zone which matches the main beam thickness measurements using the profiling phantom shown in Fig. 5(e). This indicates that the US section-thickness is small and hence the section-thickness must not introduce much error in the images captured in this setting. On the other hand, when gain is set to zero, the amount of false reflections due to the US side lobes are minimized and hence the needle tip reflections on the TRUS images are ensured to be due to intersecting the needle tips and US main beam only.

4 Conclusion

To the best of our knowledge there has been no previous work to examine the effects of imaging parameters and the US section-thickness on needle insertion depth estimation error. The US section-thickness is the combination of the both main sound-energy lobe (the main beam) and the side energy lobe. The side lobe artifacts maybe an important issue to be addressed during needle insertion procedures since they may introduce further localization errors beyond the main beam thickness artifacts.

Both the beam profile and the needle insertion experiments have provided evidence that the high-gain in US imaging setting would increase the side lobe energy of the US beam. This could result in a large elevation section-thickness profile, which directly leads to larger errors in needle insertion for TRUS-guided brachytherapy.

It is highly recommended to reduce the gain of the US imaging device to as low as practically possible to suppress the side lobe-introduced section-thickness, which would effectively minimize the needle insertion depth estimation errors (up to around 2.5 mm). The US beam profile also could help the surgeon in considering the likelihood of position errors during needle insertions.

The proposed technology is indeed tailored to brachytherapy, but the underlying principle applies to three-dimensional localization in US imagery, in general, as the US section-thickness is inherent to the modality. Side-lobes artifacts are present in every application where a needle (catheter, etc.) penetrates the US beam in the elevational direction which occurs quite ubiquitously.

Acknowledgements. This work was funded by the Natural Sciences and Engineering Research Council of Canada under the Idea to Innovation program. Gabor Fichtinger was supported as Cancer Care Ontario Research Chair. Thomas K Chen was supported as a MITACS Accelerate PhD Fellow.

References

1. Goldstein, A., Madrazo, B.L.: Slice Thickness Artifacts in Gray-Scale Ultrasound. Journal of Clinical Ultrasound 9, 365–375 (1981)
2. Skolnick, M.L.: Estimation of Beam Width in the Elevation (Section Thickness) Plane. Radiology 108, 286–288 (1991)

3. Richard, B.: Test Object for Measurement of Section Thickness at Ultrasound. Radiology 221, 279–282 (1999)
4. Chen, T.K., Thurston, A.D., Moghari, M.H., Ellis, R.E., Abolmaesumi, P.: A Real-Time Ultrasound Calibration System with Automatic Accuracy Control and Incorporation of Ultrasound Section Thickness. In: SPIE Medical Imaging (2008)
5. Peikari, M., Chen, T.K., Burdette, C., Fichtinger, G.: Section-Thickness Profiling for Brachytherapy Ultrasound Guidance. In: SPIE Medical Imaging (2011)
6. Hedrick, F.W., Hykes, D.L.: Ultrasound Physics and Instrumentation. Elsevier Mosby, Missouri (2004)
7. Laing, F.C., Kurtz, A.B.: The Importance of Ultrasonic Side-Lobe Artifacts. Radiology 145, 763–768 (1982)

Quantifying Stranded Implant Displacement Following Prostate Brachytherapy

Julio Lobo[1], Mehdi Moradi[1], Nick Chng[2], Ehsan Dehghan[3], Gabor Fichtinger[3], William J. Morris[2], and Septimiu E. Salcudean[1]

[1] Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, Canada
{juliol,moradi,tims}@ece.ubc.ca
[2] British Columbia Cancer Agency
[3] School of Computing, Queen's University

Abstract. We aim to compute radioactive stranded-implant displacement during and after prostate brachytherapy. We present the methods used to identify corresponding seeds in planned, intra-operative and post-implant patient data that enable us to compute seed displacements. A minimum cost network flow algorithm is used, on 8 patients, for needle track detection to group seeds into needles that can be matched between datasets. An iterative best line detection algorithm is used both to help with needle detection and to register the different datasets. Our results show that there was an average seed misplacement of 5.08±2.35 mm during the procedure, which then moved another 3.10±1.91 mm by the time the quality assurance CT was taken. Several directional trends in different regions of the prostate were noted and commented on.

1 Introduction

Prostate brachytherapy is an effective, minimally invasive treatment technique for men with prostate cancer[1]. It involves transperineal insertion of permanent radioactive sources, or seeds, into the prostate using needles. Many research groups have recognized the importance of monitoring the seed displacement and have used manual seed labeling, contouring or fiducial markers [2–5] to do this. We present a novel approach to automatically register different datasets and match corresponding seeds for displacement analysis.

This paper describes the computation of the misplacement, due to prostate rotation and needle deflection, of individual seeds in a stranded implant by comparing the pre-operation planned placement (*pre-plan data*) to the actual placement during surgery. The latter seed positions are computed from intra-operative fluoroscopic data [6]. We extended the study to include the quantification of seed movement from the intra-operative conditions to immediately after surgery where CT reconstruction is used to compute seed positions [7]. This movement is caused by change in patient pose from dorsal lithotomy during the procedure to horizontal recumbent during CT imaging and also from immediate inflammation. We wish to point out that although the term misplacement

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 307–314, 2011.

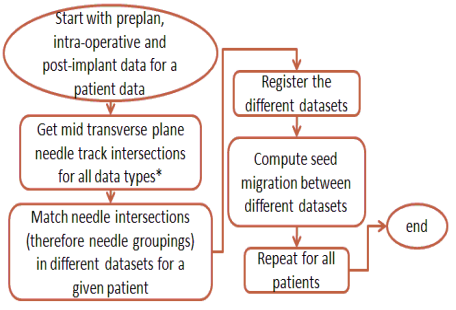


Fig. 1. Flow chart of the full process used to compute seed displacement. (*Figure 2 shows how the needle intersections are found).

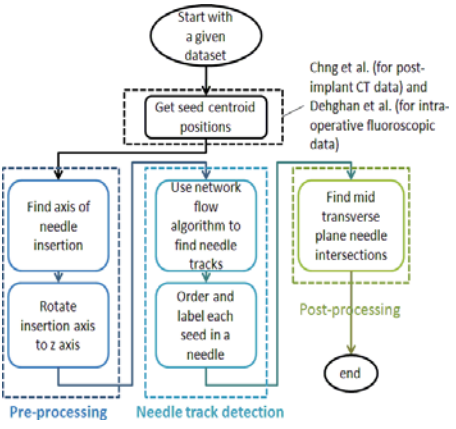


Fig. 2. Needle detection and calculation of mid transverse plane needle track intersections

may imply it, we cannot distinguish between tissue deformation and implant misplacement - analyzing this would require a separate study.

The entire process requires matching of individual seeds between these datasets to calculate the amounts of displacement. Seeds were grouped into their respective needles using a minimum cost network flow algorithm [8] (used previously by Chng *et al.* on CT data [7]). Needles were then matched so that individual seeds could be identified between datasets. To correctly compute the displacement of a given seed, a registration process was needed to put the pre-plan, intra-operative fluoroscopic and post-implant CT data into correspondence. A technical novelty for this process involved an iterative sampling algorithm to average potential needles and determine the implantation axis direction, which is not consistent between datasets.

Seed movement from pre-plan to intra-operative to post-implant have not been presented before. Seed displacement results will be presented after a description of the the methods used to label seeds in different datasets.

2 Methods

For stranded implants, seeds from the same needle stay ordered and somewhat aligned in a needle track, therefore a needle track is an intuitive class to group seeds into. Matching needle groupings between datasets simplifies the problem from matching ≈ 100 seeds to ≈ 20 needles and allows individual seed identification and therefore displacements computation.

The methods are summarized in Figure 1 and described in the next subsections. The needle detection process alone is summarized in Figure 2.

All the data used was obtained from eight consenting patients undergoing low-dose prostate brachytherapy with institutional research ethics approval.

2.1 Inter-dataset Seed Matching

Detection of the Insertion Axis. We use an iterative best line detection to find the axis of insertion of needles. Our algorithm is similar to the RANSAC (RANdom SAmple Consensus, [9]) algorithm which involves fitting lines to pairs of randomly chosen points and scoring each line. The highest scored lines are then kept as needles. Lines that have large deviations from the the z axis, which is the cranio-caudal axis in the CT, and the axis of rotation of the source-detector on the C-arm of the fluoroscope, are rejected immediately, saving unnecessary calculations. An exhaustive search on all possible pairs of points for this application is not computationally intensive so we have no random component. Pairs of points are used to fit straight lines and a Gaussian cone is used to cut-off seeds that are clearly not part of the needle.

Although the algorithm does not consistently assign seeds to needle groups correctly, the highest scoring lines are correct needle directions for a wide range of parameter values. The top eight needle directions, which are chosen to be consistent between datasets, were then averaged to find the implantation axis. This was not required for the pre-plan data which has the transducer axis already parallel to the z axis.

Needle Detection. Once the dataset is rotated so that it is aligned with the implantation axis, it is possible to use an algorithm with tighter restrictions. Because seeds from a single needle do not necessarily lie on a line, instead of looking at parameterizable lines, the seeds are viewed as nodes within a network that allow a single unit of flow through them. Each arc between two nodes carries a cost. Open source code ("matlog") written by Kay [10] was used to perform the minimization.

The cost function is made up of the same components as the scoring function iterative best line approach. Namely, these are: a measure of the angle, θ_{ij}, that the arc ij between seed i and seed j makes with the implantation axis, and a lower cost for arcs that are the length of a typical seed spacer, denoted as s_0, which is equal to integer multiples of 1 cm. The angle and spacing costs AC_{ij} and SC_{ij} are defined as follows:

$$AC_{ij} = A(e^{|\theta_{ij}|/\theta_0} - 1)^4 \tag{1}$$

$$SC_{ij} = -Be^{-s_{ij}/s_d}((\sum_{n=1}^{3} \delta(s_{ij} - ns_0)) * (e^{-s_{ij}/s_v})^2) \tag{2}$$

where s_{ij} is the spacing between node i and node j. The $*$ symbol denotes convolution. The spacing cost is defined as a Gaussian with a variance of s_v convolved with a train of delta functions to give negative peaks at integer multiples of the smallest spacing ($s_0 = 10\ mm$). A final decaying exponential with a variance of s_d is used to make the closer spacings more desirable. The spacing cost reflects the spacing seen in stranded implants.

The final objective function is:

$$C_{ij} = \gamma AC_{ij} + SC_{ij} \tag{3}$$

where γ is only used for the pre-plan data to heavily weight towards the angle.

Inter-dataset Needle Matching. After rotating the seed clusters and grouping the seeds into needles, the intersection of each needle with the most inferior transverse plane is computed. Matching is then performed using another minimum cost network flow algorithm. Each node of the network contains the x,y coordinates of an intersection as well as the number of seeds that were in the needle. Including the number of seeds per needle in each node, and the cost of a unit flow, was found to remove ambiguities caused when there are several needles close to each other.

2.2 Cluster Registration and Seed Displacement Computation

After the needles are matched, each seed position can be compared to its corresponding seed from a different data type to obtain a displacement vector. To do this the seeds clusters have to be registered so that they are in the same coordinate system. This registration was performed by translating each seed cluster so that the origin is at the centroid of all the seeds and then rotating it so that the z axis is aligned with the implantation axis. This process ensures that the centers of mass of the seed clusters are in the same place and that both the pitch and yaw of the clusters are matched. It is assumed that there is no significant differences in the rotation about the implantation axis between clusters since the bed is horizontal for both the CT and the fluoroscopic images.

An average distance was computed over all seeds from all patients for the entire volume as well as for each of the six (left, right, posterior, anterior, superior, inferior) half-sections. The average lateral, superior-inferior and anterior-posterior components of the displacement were also calculated in each section. This was done for both pre-plan to intra-operative misplacement and intra-operative to post-implant movement.

3 Results and Discussions

3.1 Inter-dataset Seed Matching Results

Iterative Best Line Detection Results. The iterative best line detection algorithm correctly assigned 79.0% to 93.2% of the seeds. 80% to 100% of the needle vectors were correctly found. In all the patients and data types, the top 8 needle directions found were correctly so the average direction gave the correct insertion axis direction.

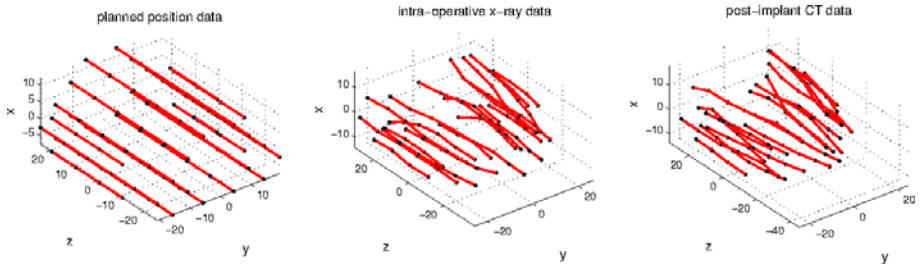


Fig. 3. Needle detection results for 1 patient for pre-plan, intra-operative and post-implant data

Minimum Cost Network Flow Needle Detection Results. Both cluster rotation and network flow calculation were tested on our 8 datasets. Figure 3 shows the results from a single patient for all three data types.

Summarizing the results, an average of 99.3% of the seeds were correctly assigned for the intra-operative data, while an average of 99.4% of the seeds were correctly assigned for the post-implant data. Once again, no pre-plan information is used in the needle detection and the algorithm takes between 1 to 2.5 seconds to rotate the cluster and find the needles. The needle detection for the pre-plan data correctly assigned all the seeds in all patients.

Inter-dataset Needle Matching Results. The few incorrectly assigned seeds in the various datasets were manually adjusted to be grouped with the correct needle. The method described earlier, achieved a correct needle matching for all the datasets, for both the pre-plan to intra-operative and intra-operative to post-implant matching.

3.2 Seed Displacement Results

With the needles correctly identified and matched in corresponding data sets for a given patient, the seed locations could be directly compared which gave us a measure of the seed displacement. As an example of a pre-plan to intra-operative comparison, Figure 4 shows the misplacement of the seeds for a single patient. Similar illustrations were computed for intra-operative to post-implant comparisons. Figure 5 shows all three sets of data for a single needle.

Average Displacement Magnitudes. The distance that every seed moved between datasets was computed for all 871 seeds from the 8 patient data sets, then averaged. The average displacement was significantly larger for the pre-plan to intra-operative case than for the intra-operative to post-implant data (5.08 mm versus 3.10 mm with $p<0.01$, $n= 871$) and both distances were within the reported localization error of 0.9 mm [6].

The average distance was computed for each half-section in the 3D space to quantify the motion in each region. Table 1 shows these distances.

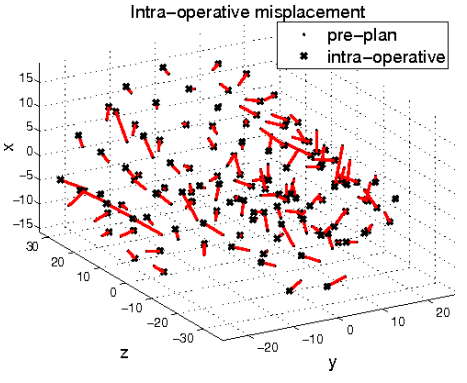


Fig. 4. Pre-plan to intra-operative seed misplacement results

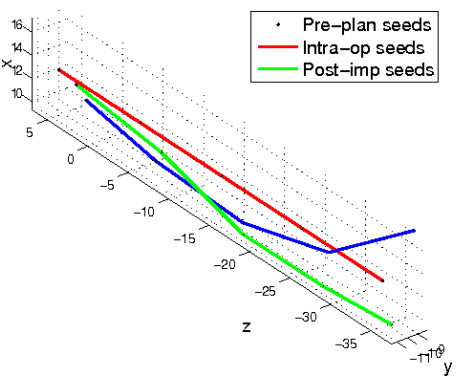


Fig. 5. Seed positions of a single needle from all three datasets

Table 1. Seed displacement results for different regions within the prostate volume

Half-section	Av. misplacement (pre to intra) /mm	Av. movement (intra to post) /mm
left	5.07±2.46	3.11±1.87
right	5.15±2.43	3.12±1.78
inferior	4.94±2.47	3.19±1.87
superior	5.29±2.42	3.04±1.78
posterior	4.63±2.34	2.97±1.61
anterior	5.60±2.55	3.26±2.04

Table 2. Displacement results for intra-operative misplacement and post-implant movement

	Intra-operative misplacement			Post-implant movement		
Half-section	lateral /mm	inf-sup /mm	ant-post /mm	lateral /mm	inf-sup /mm	ant-post /mm
left	-0.42±1.69	0.13±2.71	0.28±1.81	0.34±0.94	-0.32±1.63	0.09±1.18
right	0.44±1.93	-0.31±2.28	-0.20±1.85	-0.28±1.00	0.29±2.06	0.08±1.16
inf	-0.22±1.54	-0.27±2.66	-0.02±1.70	0.05±0.99	0.28±1.83	0.24±1.14
sup	0.24±2.11	0.09±2.34	0.10±1.99	0.01±1.05	-0.31±1.89	-0.07±1.18
post	0.34±1.16	2.02±1.30	-0.30±1.63	-0.10±1.14	0.40±1.55	-0.13±1.24
ant	-0.32±2.32	-2.19±1.32	0.38±1.98	0.16±0.86	-0.44±2.08	0.29±1.06

For both pre-plan to intra-operative and intra-operative to post-implant displacements, there were no significant differences in the amount of motion between the different half sections. All half-sections, however, had larger pre-plan to intra-operative misplacement as compared to intra-operative to post-implant movement.

Displacement Vectors. Table 2 summarizes the general displacement directions seen.

As with the scalar measurements, these directional displacements were smaller for the intra-operative to post-implant case than in the pre-plan to intra-operative case. With the calculated standard deviations, the only statistically significant displacement was the superior misplacement of the posterior seeds and the inferior misplacement of anterior seeds between the pre-plan and intra-operative conditions. No other significant trends were seen.

4 Summary and Conclusions

We have combined several techniques, including the use of an iterative best line detection algorithm and network flow algorithms, to formulate a new method to identify seeds from stranded implants in different datasets and therefore compute their displacements.

For the pre-plan to intra-operative comparison, the inferior misplacement of anterior seeds and superior misplacement of posterior seeds is in agreement with discussions with oncologists on preferred placement. Since there is a lack of implantable tissue in the anterior superior quadrant (close to the bladder), seeds are deliberately placed more inferiorly. Also, the divergence of the rectum from the prostate in the superior posterior quadrant leads to a tendency to "over-plan" this region.

The misplacement of seeds (due to needle deflection and prostate rotation) is shown to be more significant than the movement immediately after surgery. This agrees with work done by Wan et al. to evaluate needle deflection [11] and by Lagerburg et al. who evaluate prostate rotation during the insertion of needles [12]. The results also emphasize the seed placement process as a research area that requires more attention, in comparison to immediate post-implant seed movement, in future efforts to improve prostate brachytherapy.

The algorithms explained in this paper have been described for pre-plan to intra-operative fluoroscopic to post-implant CT-data. However, they can also be used to compare seed positions over several days after a surgery to further monitor inflammation. Another application of the developed algorithms is in real-time dosimetry. Complete ultrasound-based brachytherapy seed detection is a notoriously challenging problem. However, Moradi *et al.* have shown that given a complete reconstruction of the fluoroscopic data and identification of the seeds, registration can be performed with an incomplete reconstruction of the seed cloud in ultrasound data [13]. Therefore, the complete seed identification (or labeling) algorithm described here, could make it possible to perform intra-operative dosimetry.

Acknowledgments. The work was supported by NSERC and CIHR. Gabor Fichtinger was supported as a Cancer Care Ontario Research Chair.

References

1. Morris, W., Keyes, M., Palma, D., Spadinger, I., McKenzie, M., Agranovich, A., Pickles, T., Liu, M., Kwan, W., Wu, J.: Population-based study of biochemical and survival outcomes after permanent 125I brachytherapy for low- and intermediate-risk prostate cancer. Urology 73(4), 860–865 (2009)
2. Roberson, P.L., Vrinda, N., Daniel, L.M., Raymond, J.W., McLaughlin, P.W.: Effects of seed migration on post-implant dosimetry of prostate brachytherapy. Medical Physics 24, 251–257 (1997)
3. Pinkawa, M., Gagel, B., Asadpour, B., Piroth, M., Klotz, J., Borchers, H., Jakse, G., Eble, M.: Seed displacements after permanent brachytherapy for prostate cancer in dependence on the prostate level. Strahlentherapie und Onkologie 184, 520–525 (2008) 10.1007/s00066-008-1886-3
4. Pinkawa, M., Asadpour, B., Piroth, M.D., Gagel, B., Klotz, J., Fischedick, K., Borchers, H., Jakse, G., Eble, M.J.: Rectal dosimetry following prostate brachytherapy with stranded seeds - comparison of transrectal ultrasound intra-operative planning (day 0) and computed tomography-postplanning (day 1 vs. day 30) with special focus on sources placed close to the rectal wall. Radiotherapy and Oncology 91(2), 207–212 (2009)
5. Usmani, N., Chng, N., Spadinger, I., Morris, W.J.: Quantification of migration of 125iodine rapidstrands in prostate brachytherapy implants in different regions of the prostate. In: Genitourinary Cancers Symposium, vol. 96 (2010)
6. Dehghan, E., Lee, J., Moradi, M., Wen, X., Fichtinger, G., Salcudean, S.E.: Prostate brachytherapy seed reconstruction using C-arm rotation measurement and motion compensation. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6361, pp. 283–290. Springer, Heidelberg (2010)
7. Chng, N., Spadinger, I., Morris, W., Usmani, N., Salcudean, S.: Prostate brachytherapy postimplant dosimetry: Automatic plan reconstruction of stranded implants. Brachytherapy (2010)
8. Ahuja, R., Magnanti, T., Orlin, J.: Network Flows: Theory, Algorithms, and Applications. Prentice Hall, Englewood Cliffs (1993)
9. Fischler, M.A., Bolles, R.C.: Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Comm. of the ACM, 381–395 (1981)
10. Kay, M.G., Parlikad, A.N.: Material flow analysis of public logistics networks. Progress in Material Handling Research, 205–218 (2002)
11. Wan, G., Wei, Z., Gardi, L., Downey, D.B., Fenster, A.: Brachytherapy needle deflection evaluation and correction. Medical Physics 32(4), 902–909 (2005)
12. Lagerburg, V., Moerland, M.A., Lagendijk, J.J., Battermann, J.J.: Measurement of prostate rotation during insertion of needles for brachytherapy. Radiotherapy and Oncology 77(3), 318–323 (2005)
13. Moradi, M., Wen, X., Dehghan, E., Lobo, J., Morris, W.J., Salcudean, S.E.: Needle path detection for brachytherapy dosimetry based on lateral power imaging and template matching. In: IEEE Ultrasonics Symposium, pp. 586–589 (2009)

Preconditioner-Based Contact Response and Application to Cataract Surgery

Hadrien Courtecuisse, Jérémie Allard, Christian Duriez, and Stéphane Cotin

SHAMAN Group, INRIA Lille North Europe

Abstract. In this paper we introduce a new method to compute, in real-time, the physical behavior of several colliding soft-tissues in a surgical simulation. The numerical approach is based on finite element modeling and allows for a fast update of a large number of tetrahedral elements. The speed-up is obtained by the use of a specific preconditioner that is updated at low frequency. The preconditioning enables an optimized computation of both large deformations and precise contact response. Moreover, homogeneous and inhomogeneous tissues are simulated with the same accuracy. Finally, we illustrate our method in a simulation of one step in a cataract surgery procedure, which require to handle contacts with non homogeneous objects precisely.

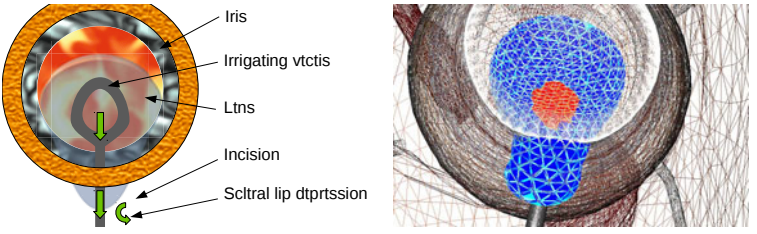
1 Introduction

Several methods have been presented for real-time bio-mechanical simulations of contacting soft-tissues. However, this work is motivated by a training simulator for cataract surgery, whose needs can not be addressed with existing methods. Thus, we first present the context and the motivations for this work before highlighting simulation requirements and the contributions of our method.

Context: In developed countries, when the crystalline lens is being clouded over by a cataract, a therapy based on *phacoemulsification* is commonly used. Training simulation of this procedure has been studied [3,4], and a commercial solution is provided by VRmagic®. Another surgical procedure known as Manual Small Incision Cataract Surgery (MSICS), requires low technology with almost equivalent results when performed by a well-trained specialist. A recent report [7] shows that there is a great need of training for this surgery, to face the huge number of people that need care. This work provides the first steps toward a simulator that could help solve this training bottleneck.

Unlike during phacoemulsification, MSICS consists in pulling out the crystalline lens as a single piece through a small incision (see Fig. 1). This step generates large deformations on both lens and eyeball. Moreover, for older patients, the center of the lens can be much stiffer than the outer part. This inhomogeneity must be taken into account, as it impacts the successful completion of the surgery. Indeed, it changes the mechanical behaviors and may require adapting the size of the incision. A training system for MSICS necessitates an accurate modeling of these soft-tissues in a simulation executed in real-time.

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 315–322, 2011.



(a) Extraction of the nucleous (b) Simulated procedure

Fig. 1. Nucleus extraction with an irrigating vectis. (a) motion of the tool, (b) simulated lens. Elements at the center (in red) are stiffer than at the periphery (blue).

Contributions: In this context of tradeoff between accuracy and computation time, we need: (1) to account for precise contact response in highly constrained cases and for strong stiffness inhomogeneities. We note that these inhomogeneities create ill-conditioned systems that are more difficult to solve. (2) to use a large number of nodes and elements in the finite element formulation, in order to maintain the realism of the deformations.

The contribution of this paper is to answer these needs (which probably exist for other surgical procedures) in a real-time simulation. The method is based on an implicit integration scheme (to manage the large range of stiffnesses while keeping large time steps) and uses a preconditioner that is updated at low frequency. This preconditioner provides almost an exact solution to the deformation solver, allowing its fast convergence. Moreover, it provides a very good estimation of the mechanical coupling between the contacting nodes. This allows for a precise constraint solving process based on accurate contact laws.

2 Related Works

Many previous works involve the simulation of soft body in contacts. We will concentrate here on aspects related to the proposed strategies to solve the mechanical and contact equations once they have been expressed.

The core mechanical equations are based on Newton's second law and the Finite Element Method (FEM). We rely on a linearized implicit integration scheme [2], which is more stable when sudden contacts occur. Thus we update the next positions and velocities with the following equations :

$$\mathbf{M}\mathbf{a}_{t+h} = \mathbf{f}(\mathbf{x}_{t+h}, \mathbf{v}_{t+h}) \qquad \mathbf{v}_{t+h} = \mathbf{v}_t + h\mathbf{a}_{t+h} \qquad \mathbf{x}_{t+h} = \mathbf{x}_t + h\mathbf{v}_{t+h} \tag{1}$$

$$\underbrace{\left(\mathbf{M} + h\mathbf{B} + h^2\mathbf{K}\right)}_{\mathbf{A}}\mathbf{a}_{t+h} = \underbrace{\mathbf{f}(\mathbf{x}_t, \mathbf{v}_t) + h\mathbf{K}\mathbf{v}_t}_{\mathbf{b}} \tag{2}$$

where $\mathbf{M}$ is the *mass* matrix, $\mathbf{B}$ the *damping* matrix and $\mathbf{K}$ the *stiffness* matrix.

Eq. (2) must be solved at each time-step, but continuously changes due to material and geometrical non-linearities. As the system matrix is positive semi-definite, a popular algorithm to efficiently solve this problem is the Conjugate

Gradient (CG) iterative solver. However, it suffers from convergence issues for ill-conditioned matrices, which can appear for inhomogeneous materials or finite element meshes of non optimal quality. Preconditioning techniques consists in computing an approximation $\mathbf{P}$ of matrix $\mathbf{A}$ which is easier to invert. Then, this preconditioner can be used to solve (2) by relying on $\mathbf{P^{-1}A}$, which is better conditioned and thus converge to an adequate solution in fewer iterations. However, computing a preconditioner adds two overheads: first to invert (or factorize) the preconditioner itself, then to apply it at each CG iteration. Several preconditioners can be used, from simple diagonal matrices [2] to costly but precise incomplete Cholesky factorizations.

When contacts occur, they must be taken into account in the above equation system. A common solution consists in using a penalty method, which handle contacts by adding a contact force $\mathbf{f} = k\delta\mathbf{n}$ at each contact point, where δ is a measure of the interpenetration, $\mathbf{n}$ is the contact normal and k is a stiffness factor. This stiffness is difficult to determine, as it must be tuned according to the stiffness of the object in contact. This becomes particularly an issue when dealing with inhomogeneous objects. Stiff penalty forces can either be considered as explicitly-integrated external forces, which can introduce instabilities and would require lowering the time-steps. Alternatively, they can be placed in the same implicit equations as the internal forces, creating a large system that can be difficult to solve.

Other methods rely on Lagrange multiplier [8,10,6] to compute contact forces such that all intersections are removed at the end of each time step. The core computation of the algorithm involves solving a Linear Complementary Problem (LCP) deriving from Signorini's law :

$$\boldsymbol{\delta} = \mathbf{HCH}^T\boldsymbol{\lambda} + \boldsymbol{\delta}_0 \qquad \text{with} \qquad 0 \leq \boldsymbol{\delta} \perp \boldsymbol{\lambda} \geq 0 \tag{3}$$

where $\mathbf{H}$ is the Jacobian of the contacts, and $\mathbf{C}$ is the *compliance* matrix, λ is the contact force, $\boldsymbol{\delta}_0$ and $\boldsymbol{\delta}$ are the measure of interpenetrations before and after collision response. This equation means that, if λ is positive at the end of the time step, then δ must be equal to zero, and vice versa.

For explicit methods $\mathbf{C}$ is the inverse of the (often diagonalized) mass matrix, whereas for implicit schemes it also involves damping and stiffness, i.e. $\mathbf{C} = (\frac{1}{h^2}\mathbf{M} + \frac{1}{h}\mathbf{B} + \mathbf{K})^{-1}$ for the scheme used in (2). This matrix changes at every time steps, and its computation can be prohibitive for large deformable meshes. It is possible to use an approximation of the compliance, because an approximated local deformation in response to a collision is visually acceptable. The most extreme approximation is to only consider the mass, as for explicit schemes. When it is used, contacts are corrected without any mechanical coupling between nodes. A more accurate method has been proposed in [11], inspired by the co-rotational formulation used in FEM to remove non-linearities introduced by rotations. $\mathbf{C}$ can be precomputed from the rest configuration and updated at each time step based on a local estimation of the rotations. However, this approximation can become inaccurate for large deformations. Moreover, it requires storing a dense matrix, with $9n^2$ values where n is the number of vertices of the object, therefore preventing its application on detailed meshes.

3 Inhomogeneous Deformable Model on GPU

Soft tissue deformations are modeled using a geometrically non-linear elastic formulation and a Finite Element Method (co-rotational model introduced in [9]). The local rotation of each element is estimated in order remove the influence of geometrical nonlinearities, allowing to account for large deformations. To compute the deformation in real-time with detailed meshes (more than 10,000 tetrahedral elements), we implemented a CG solver on the GPU [1].

Recently, we introduced a new approach [5] to maintain a good approximation of the inverse of the system matrix during the simulation. This method exploits time coherency to update an exact factorization of the system matrix within a separated loop computed at a lower frequency. It only requires a few simulation steps to provide a new factorization, thus constantly providing a good approximation usable as a preconditioner. It is further improved by estimating rotations between the current position and the state used for the last factorization, similar to the co-rotational formulation. Even if the preconditioner is applied on the CPU, the GPU can still be used for the remaining operations. Finally, as it uses a sparse factorization, it supports simulations with a large number of elements.

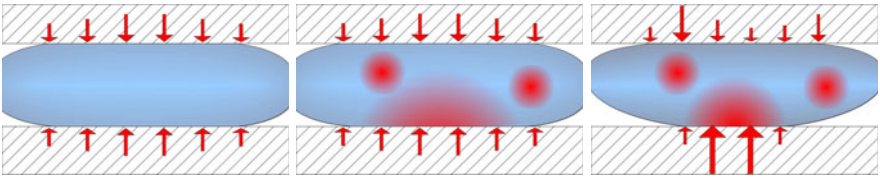
As a consequence the above method significantly reduces the number of iterations needed for the CG algorithm to converge. However, the paper only considered simulations involving a single object without any contacts. In this paper, we extend this idea to the computation of contact responses, as is detailed in the next section, and apply this new technique in the field of medical simulation. In the context of the cataract simulation, two deformable tissues need to be simulated: the eye lens and sclera (Fig. 1) which both undergo large deformations combined with multiple contacts during surgery. Furthermore, the crystalline lens is by nature inhomogeneous, with a stiff kernel and softer boundaries, while the meshes of these structures contain elements of potentially very different size (in particular near the incision located in the sclera). All of these characteristics lead to a very poor conditioning of the system matrix, and it is obvious that in this case, preconditioning techniques could greatly improve the converge rate of the solver. For this application, we found the $\mathbf{LDL}^{\mathrm{T}}$ factorization to be more stable than the $\mathbf{LL}^{\mathrm{T}}$ Cholesky factorization. Therefore, the complete expression of the preconditioner we use is:

$$\mathbf{P} = \mathbf{R}^{\mathrm{T}}_{t \to t-\Delta t} \ \mathbf{L_{t-\Delta t} D_{t-\Delta t} L^{T}_{t-\Delta t}} \ \mathbf{R}_{t \to t-\Delta t} \tag{4}$$

where $\mathbf{R_{t \to t-\Delta t}}$ is the current local rotations matrix, and $\mathbf{L_{t-\Delta t} D_{t-\Delta t} L^{T}_{t-\Delta t}}$ the most recent factorization.

4 Preconditioner for Contact Response

Local mechanical coupling must be taken into account to compute accurate collision response (see Fig 2). Indeed, while the constraint-based formulation will ensure a contact-free configuration at the end of each time-step in all cases, the local deformation near contacts obtained without the correct compliance (Fig. 2(b))



(a) Homogeneous object (b) Inhomogeneous object without coupling (c) Inhomogeneous object with coupling

Fig. 2. Force repartition for homogeneous and inhomogeneous objects on which collisions are solved with and without mechanical coupling. Stiffer part are shown in red.

is invalid. While this difference would be reduced after several time-steps, using a more accurate compliance produces better intermediate configurations and, more importantly for real-time simulations, allow larger time-steps.

To solve this issue, we propose to extend the work introduced in [11], using the preconditioner as compliance matrix instead of a precomputed dense inverse matrix. Indeed, the compliance matrix is only a scaled version of the inverse system matrix: $\mathbf{C} = h^2\mathbf{A^{-1}}$. Moreover, we showed in section 3 that we maintain a good approximation of $\mathbf{A^{-1}}$ by updating the preconditioner at low frequency. Thus, if $\mathbf{P}$ remains a good approximation of the system, it can be re-used to build the compliance matrix. That way, we guarantee a contact-free configuration at the end of the time step with almost the exact mechanical coupling of elements taken into account. To build the LCP matrix, we combine eq. (3) and (4) to compute for each object:

$$\begin{aligned} \mathbf{H}\mathbf{C}\mathbf{H}^{\mathrm{T}} &= h^2\,\mathbf{H}\mathbf{A}^{-1}\mathbf{H}^{\mathrm{T}} \approx\; h^2\,\mathbf{H}\left(\mathbf{R}\mathbf{L}\mathbf{D}\mathbf{L}^{\mathrm{T}}\mathbf{R}^{\mathrm{T}}\right)^{-1}\mathbf{H}^{\mathrm{T}} \\ &\approx h^2\,\mathbf{J}\left(\mathbf{L}\mathbf{D}\mathbf{L}^{\mathrm{T}}\right)^{-1}\mathbf{J}^{\mathrm{T}} \qquad \text{with} \quad \mathbf{J} = \mathbf{H}\mathbf{R} \end{aligned} \tag{5}$$

This computation can be implemented in three steps. The first step consists in applying the local rotations since the last update to $\mathbf{H}$. This operation is inexpensive because it is done by computing the product of a block-diagonal matrix $\mathbf{R}$ with a sparse matrix into matrix $\mathbf{J}$ with the same sparsity structure. Next, we compute the product of $\mathbf{J}^{\mathrm{T}}$ with the inverse of the factorization. This is achieved by computing rows independently, each requiring two sparse triangular solves (STS) using one row of $\mathbf{J}$. Finally, the resulting matrix is multiplied by matrix $\mathbf{J}$ to obtain the final contribution to the LCP matrix.

The number of STS in step 2 is proportional to the number of contacts, which can become very large in some cases. However, as each row is independent, we proposeto implement them on GPU, by computing several STS in parallel, reducing the level of parallelism that must be achieved within each STS. This maps nicely to the two-level SIMD architecture of todays GPU where synchronizations within a group of cores is fast, whereas global synchronization over multiple groups is much more prohibitive. Such synchronizations are necessary to guarantee the respect of dependencies within the computation, and would be

difficult to implement efficiently if only a single STS was computed on all GPU computing cores.

Matrices **H**, **J** and **L** can be stored in *compressed sparse row* (CSR) format to save memory. However in the GPU implementation it is more efficient to use a dense representation for **J**, such that the elements of the right-hand vectors for each solve can be accessed directly. Another consequence is that the final matrix product then involves two dense matrices, and can therefore be implemented using a standard BLAS library, such as CUBLAS for NVIDIA® GPUs. While the LCP matrix is computed on GPU, it is then solved on CPU using Gauss-Siedel algorithm.. The size of this matrix only depends on the number of contacts, which is often much smaller that the number of mechanical vertices, so this transfer and solve is much faster that the previous steps in most simulations, and the gain to implement them on GPU would be negligible.

5 Results

Evaluation: We measured the error introduced by using an approximation to the compliance, compared to an exact factorization at each time step. We produced a simulation involving a non-homogeneous disc pushed by a sphere through a small hole. The difference of vertex positions was compared using Root Mean Square error (see Fig. 3), when using only the diagonal matrix and our method with several update frequencies, with *no update* being equivalent to [11], and *async* corresponding to our contribution where the preconditioner is updated as soon as possible using another thread. Results show that using inaccurate approximation does not lead to the correct behavior. Using a diagonal compliance does not consider the coupling between vertices within each time-step and is stable only for very small steps, relying on the mechanical computations to propagate contact forces. With the proposed method, we can simulate such scenario in real time if the preconditioner can be updated sufficiently frequently. Indeed, when the factorization is updated more often, the obtained shape and

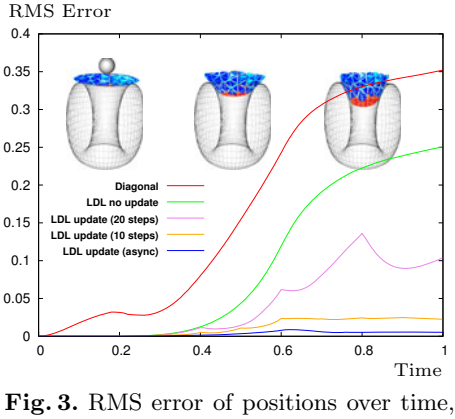


Fig. 3. RMS error of positions over time, introduced by several approximation as compared to an exact factorization

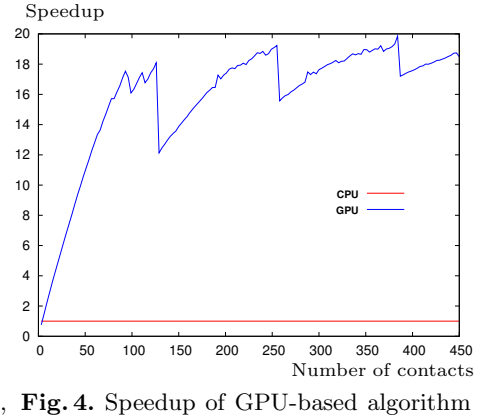


Fig. 4. Speedup of GPU-based algorithm to build the compliance matrix according to the number of contacts

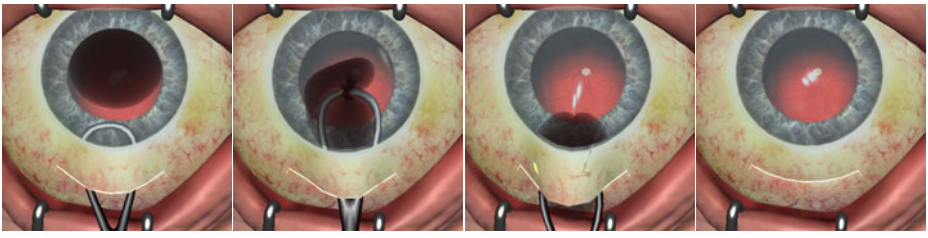


Fig. 5. Simulation of the extraction of the eye lens during MSICS

behavior is increasingly accurate. In this example, the multi threaded version updates the preconditioner on average every 3 steps, allowing to simulate almost the exact behavior.

GPU Implementation: We measured the computation time (see Fig. 4) required to build the compliance matrix with our GPU-based algorithm compared to a sequential CPU implementation. As expected, the computation time to obtain a single solve with our GPU algorithm (21.90 ms) is larger than the CPU (2.44 ms) version. Thus, we maintain the application of the preconditioner during the mechanical CG solver on the CPU (as each iteration only requires one solve). However, the computation time for solving multiple contacts remains almost constant with the GPU algorithm, which allows to quickly take advantage of the graphic processor. Indeed, until 130 contacts the computation units of the GPU are not fully exploited, and each STS is computed in parallel. Beyond this number of contacts, some of the computation units will compute several STS successively. However, the GPU is able to overlay waiting times, due to synchronizations and read/write operations in memory, with computations for another STS. Thus, solving 260 contacts is only 1.5 times slower than 130.

MSICS Simulation: Finally, we show in Fig. 5 our simulation of the extraction of an opacified lens during a MSICS simulation. The lens is modeled with $1,113$ vertices and $4,862$ tetrahedra, whereas the eye contains $1,249$ vertices and $3,734$ tetrahedra. The center of the lens is five times stiffer than the periphery. For both structures, we used the LDL-based preconditioner, which was updated on average every 4 time steps for the eye and 3 time steps for the lens. With this preconditioner, the CG required an average of 8.5 iterations to converge.

The lens is removed with the help of the deformations of the sclera and friction between the tool and the lens. The surgical instrument is controlled by the user with an haptic device, with the method introduced in [12]. As we model the compliance of the objects in contact with the instrument, we are able to compute and transmit feedback forces to the device.

The simulation runs between 15 and 50 FPS depending on the number of contacts, which varies between 50 and 90 during the extraction. The computation within a single time is split between 31.5% for the preconditioned CG, 47.40% for the GPU-based assembly of the compliance matrix, and less than 22% for the other computations, including collision detection and LCP solver.

All the previous results were obtained on a quad-core Intel® Core™ i7 3.07 GHz CPU with a Nvidia® GeForce® GTX 580 GPU.

6 Conclusion

In this paper, we have presented a novel method for solving accurate contacts response between soft tissues which is based on the use of a periodically updated preconditioner. The proposed method allows to simulate a large number of elements and supports both homogeneous and inhomogeneous objects with a similar accuracy, while taking into account the mechanical coupling between contacts. We applied our method to a MSICS simulation, which presents very constrained situations during the eye lens extraction. Although this method was illustrated in a cataract surgery, it can be applicable to any simulations with similar requirements. As future works, we would like to simulate the complete simulation of MSICS, including the incision and capsulorhexis, as well as handling the influence of fluid pressure within the eye.

References

1. Allard, J., Courtecuisse, H., Faure, F.: Implicit FEM solver on GPU for interactive deformation simulation. In: GPU Computing Gems Jade Edition, ch.21, Elsevier, Amsterdam (2011)
2. Baraff, D., Witkin, A.: Large steps in cloth simulation. In: Proc. of SIGGRAPH 1998, pp. 43–54. ACM, New York (1998)
3. Choi, K.S., Soo, S., Chung, F.L.: A virtual training simulator for learning cataract surgery with phacoemulsification. Comput. Biol. Med. 39(11), 1020–1031 (2009)
4. Comas, O., Cotin, S., Duriez, C.: A shell model for real-time simulation of intraocular implant deployment. In: Bello, F., Cotin, S. (eds.) ISBMS 2010. LNCS, vol. 5958, pp. 160–170. Springer, Heidelberg (2010)
5. Courtecuisse, H., Allard, J., Duriez, C., Cotin, S.: Asynchronous preconditioners for efficient solving of non-linear deformations. In: Erleben, K., Bender, J., Teschner, M. (eds.) VRIPHYS, pp. 59–68. Eurographics Association (2010)
6. Galoppo, N., Otaduy, M.A., Mecklenburg, P., Gross, M., Lin, M.: Fast simulation of deformable models in contact using dynamic deformation textures. In: Cani, M.-P., O'Brien, J. (eds.) SCA, pp. 73–82. Eurographics Association (2006)
7. HelpMeSee, http://helpmesee.org/about/statement-of-purpose/
8. Lenoir, J., Grisoni, L., Meseure, P., Rémion, Y., Chaillou, C.: Smooth constraints for spline variational modeling. In: GRAPHITE 2004, pp. 58–64. ACM, New York (2004)
9. Nesme, M., Payan, Y., Faure, F.: Efficient, physically plausible finite elements. In: Eurographics (short papers), pp. 77–80. Eurographics Association (2005)
10. Popescu, D., Compton, M.: A model for efficient and accurate interaction with elastic objects in haptic virtual environments. In: GRAPHITE 2003, pp. 245–250. ACM, New York (2003)
11. Saupin, G., Duriez, C., Cotin, S., Grisoni, L.: Efficient contact modeling using compliance warping. In: Computer Graphics International Conference (2008)
12. Saupin, G., Duriez, C., Cotin, S.: Contact model for haptic medical simulations. In: Bello, F., Edwards, E. (eds.) ISBMS 2008. LNCS, vol. 5104, pp. 157–165. Springer, Heidelberg (2008)

Layered Surface Fluid Simulation for Surgical Training

Louis Borgeat, Philippe Massicotte, Guillaume Poirier, and Guy Godin

National Research Council of Canada
first.last@nrc-cnrc.gc.ca

Abstract. We present a novel approach to fluid simulation over complex dynamic geometry designed for the specific context of virtual surgery simulation. The method combines a surface-based fluid simulation model with a multi-layer depth peeling representation to allow realistic yet efficient simulation of bleeding on complex surfaces undergoing geometry and topology modifications. Our implementation allows for fast fluid propagation and accumulation over the entire scene, and runs on the GPU at a constant low cost that is independent of the amount of blood in the scene. The proposed bleeding simulation is integrated in a complete simulator for brain tumor resection, where trainees have to manage blood aspiration and tissue/vessel cauterization while they perform virtual surgery tasks.

Keywords: visual simulation, GPU, fluid simulation, bleeding, cauterization, aspiration, depth peeling, surgery, neurosurgery.

1 Introduction

The key challenge in virtual surgery simulation lies in the large amount of computer resources required for real-time interactive simulation of complex physics and for realistic haptic and visual feedback. This paper contributes a technique for the efficient integration of a proven and low-cost fluid simulation approach to the context of bleeding simulation for surgery, where the fluid must adapt to the constantly transforming geometry associated with the tissue simulation.

The research presented in this paper was performed in the context of the development of a complete patient-specific brain tumor resection simulator, shown in Figure 1. The simulator provides haptic feedback for two hands/tools, finite element (FE) tissue modeling and simulation with topological adaptation during resection, and realistic visual feedback including high resolution texturing and lighting, smooth cast shadows, depth of field, lens distortion, and bleeding. The simulator is currently being evaluated in hospitals for resident training.

Significant research efforts have already been targeted at fluid simulation in the context of bleeding for surgery training. Andersson [1] surveys the various fluid effects required by surgery simulation. Our specific requirement, established with clinical partners, was for a model supporting fast and realistic surface blood flow and accumulation over the entire simulated surface of the surgical zone.

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 323–330, 2011.

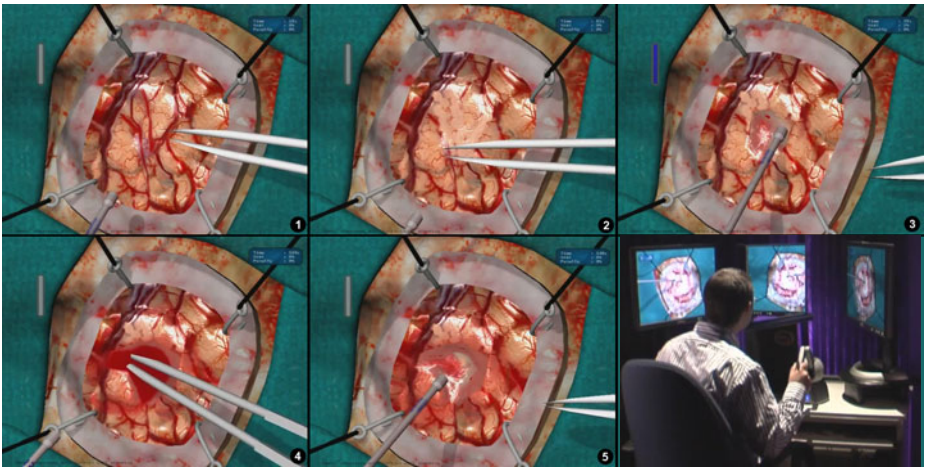


Fig. 1. Brain tumor resection with the Neurotouch simulator. 1. Initial brain surface. 2. After surface vessels cauterization with virtual bipolar tool. 3. Tissue aspiration. 4. Blood accumulation and cauterization. 5. Final blood aspiration, bleeding has been stopped. 6. Simulator prototype with haptic tools and stereoscopic display.

Among particle-based fluid simulation techniques, many recent contributions expand on the Smoothed Particle Hydrodynamics method [7, 10, 11], but such techniques were too slow for our application. Most other solutions are grid-based approaches in either 3D or 2D. In full 3D, good results have been achieved with 3D simulations using cubic interpolated propagation [13] that allows the use of a coarser simulation grid. Among 2D techniques, Basdogan [2] presents a real-time model of blood flow over soft tissues that projects a two-dimensional grid placed over the region of interest of the geometric tissue model. Kerwin [5] uses a two-dimensional Eulerian fluid simulation performed on the GPU which is integrated with a volume renderer. Numerous works build on height field (water column) methods where fluid equations are implemented in two dimensions using hydrostatic pressure pipe models [4, 9, 12]. These methods allow for fast computation, wave-like behavior, net fluid transport, and dynamic boundary conditions, but due to their inherent 2D nature their application scope is limited. Such approaches are also used with physically-based erosion simulation models [15, 6]. For gaming, simpler models have also been proposed for fluid simulation on textures [8]. But such methods usually rely on surface parameterizations that could not easily be maintained under major surface topology changes.

In this paper, we adapt the class of water column surface models to the case of complex and deforming/transforming scene geometries by applying them over a layered depth image version of the geometric model produced using a depth peeling approach [3]. The mesh is rendered at each frame into a stack of height maps. Then, an extended version of the GPU implementation of a grid-based fluid model proposed by Mei [6] is used to propagate blood on the surface. The

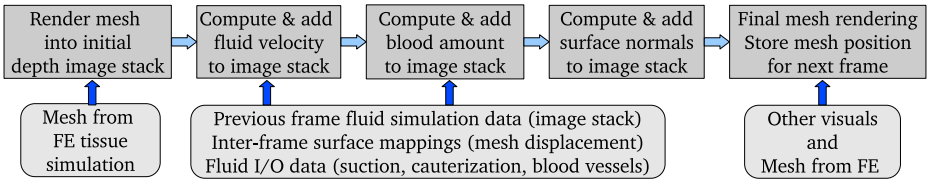


Fig. 2. Top: Processing sequence for a given frame. Bottom: input data at each step.

resulting fluid images are used when rendering the deforming geometry to add proper color, shading and elevation to the mesh. This approach brings many benefits:

- By working on images on the GPU, we achieve a fast and low-cost simulation, thus leaving most resources available for other tasks.
- Independence from the simulated mesh: by working in image space, we do not need to compute contact with the geometry, but only to be able to render it. We simply need to receive vertex offsets at each frame and a list of changed/new vertices when the topology is modified. In fact, the mesh's physical deformation could actually be simulated on a different machine.
- There is no need for a surface parameterization.
- Performance is independent of the amount of blood in the scene.

The remainder of the paper describes how a grid-based fluid model can be efficiently mapped onto a deforming and topology-changing mesh, using a water-column model for our application. The integration of the fluid simulation with the rendering as well as the performance of the method are also discussed.

2 Bleeding Simulation

In order to run an image-based fluid simulation on the deforming surface mesh computed for haptics and tissue simulation, we repeat the following sequence for each frame (Fig. 2). The mesh is first converted on the GPU into a stack of depth images rendered from a view direction aligned with gravity. We then run the modified column-based simulation on those multiple interconnected depth layers, fetching information from the image stack of the previous frame as the base for the simulation step. We finally render the mesh with the bleeding by using the simulation layers as input texture. During this process, one important challenge is to maintain a mapping between consecutive frames so that inter-frame fluid propagation takes into account mesh displacement and transformation.

Mesh to Depth Maps: The conversion of the mesh into a depth image stack is achieved by a process called depth peeling [3]: the polygonal mesh is first rendered at a chosen resolution under an orthographic view aligned with gravity and closely encompassing the entire scene. Depth/height rendered from that

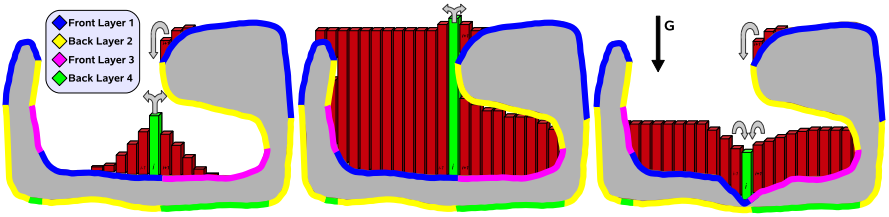


Fig. 3. Slice view of the depth peeling process and blood propagation for a simple geometric model with an overhang. Color denotes the depth layer on which a portion of the surface is represented. We illustrate 3 complex cases of blood propagation at the junction between different layers: (a) the central (green) cell gives to neighbors on the mesh, but located on different layers; (b) because of accumulation, the center cell nows gives blood to a different (higher) layer; (c) the cell receives blood from two different layers on the right

viewpoint is stored in a floating-point GPU buffer. This process is repeated in a sequence of additional buffers, but each time all the geometry located above or at the previous height map is culled at the fragment/pixel level. We therefore end up with a decomposition of the mesh into a stack of occluding depth layers (color-coded in Fig. 3). If the 3D model is closed and not self-intersecting, the peeling process will naturally produce an alternation of front and back facing layers. In our implementation, we use the front (upward) facing layers to run the actual simulation, and keep the information of the back facing layers to insure that blood cannot accumulate beyond the upper surfaces of an enclosed area. To accelerate computations, we pack front facing layers in a single texture buffer and back facing layers in another.

Fluid Simulation: The core of the fluid simulation is similar to the work of Mei [6], but with the added difficulty that fluid must propagate properly between the different layers of the stack representing the model at a given frame and, as previously stated, between the layers associated with consecutive frames as they deform. The equations are also adapted to our application context, but the physics remains the same: cells exchange fluid with their 4 neighbors through virtual pipes, and a flow velocity, or flux, is updated at each frame based on hydrostatic pressure in the virtual pipe between the neighboring cells. Fluid propagation is computed in two rendering steps. After generating new depth layers, a first step computes a set of outflow flux (velocity) textures, setting the amount of outgoing fluid in all 4 directions and properly identifying the depth layer where the fluid must be transferred for each direction, taking into account the amount of fluid available in the cell, and the remaining available height in the target cell. Initial blood amount and flux/velocity are obtained from the last frame textures as described in the next section. Flux equations are controlled by

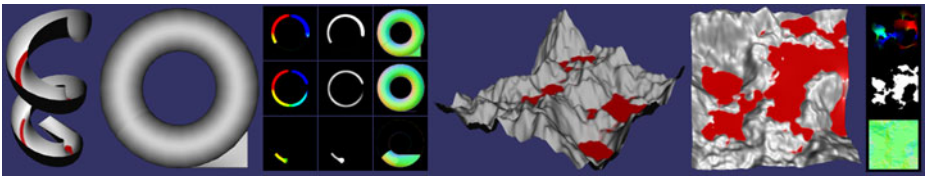


Fig. 4. Two simple models to illustrate the GPU buffers used for the fluid simulation and for the depth peeling process.Left: Depth peeling and partial simulation on a helical slide model. Buffers from left to right: Blood velocity, amount of blood, and surface normals. Right: Simulation on simple terrain model with simulation GPU buffers.

application-specific parameters. For example, in the blood flux computation, we set a minimum amount of blood in a cell before it can be transferred to another one, in order to force trainees to clean tissue surfaces with the suction device.

Figure 3 illustrates different cases of inter-layer fluid exchange within a frame. As can be seen in the figure, continuous surfaces can be located on different layers of the depth image stack. When computing output flux for a cell, one target layer is selected separately in each 4 directions. However, a cell can receive fluid from the same direction from multiple neighbors located in different layers (Fig. 3(c)). The case in Figure 3(b) illustrates that, as blood accumulates, it may be sent to different layers even if the geometry does not change. When computing target layers for fluid output in one direction, we compare the height of the cell (including blood) to the height of neighboring cells (also with blood) of all layers in that direction. By integrating the maximum height (back faces) information in the comparison process, the correct target can be deterministically selected.

The second fluid simulation step computes the resulting amount of blood in a cell by gathering the flux value for each cell into a set of textures representing fluid amounts, and then modifying that value based on new bleeding and aspiration that may occur within that frame. The amount of new blood generated for a given layer cell is obtained from other components of the simulation. In our implementation, the FE volume structure from which the rendered surface mesh is extracted is associated through a fixed parameterization to a set of 3D volumetric textures. A high resolution texture is derived from medical imaging enhanced with synthetic data and provides vascularization information at the given location. Another 3D texture storing cauterization levels is updated locally at each frame based on the location of active cauterization tools. We also tag vertices in the mesh based on their creation or modification time, which correspond to the time of the last tissue cutting. All this information is combined to compute the amount of new blood at each cell. Information on the location and orientation of the blood aspiration tools is passed to the same GPU program to apply suction. Figure 4 shows the content of the flux and fluid amount textures for two simple illustrative models.

Inter-Frame Fluid Propagation: The depth image layers corresponding to consecutive simulation frames have different shapes since the rendered surface

mesh undergoes continuous transformation. We therefore must establish a mapping between these stacks to propagate the fluid and its velocity from frame to frame. To achieve this, each vertex in the geometric model retains its coordinates in the previous frame as a vertex attribute. After the peeling, we render all the vertices into a dedicated floating-point texture to store that information for the next frame.

During the fluid simulation step, a texture position and depth in the previous frame's depth image stack is computed for each cell to allow texture fetches in the previous frame data. This position is produced on the GPU by interpolating the spatial position of the surrounding vertices in the last frame. Since the texels are generally not aligned between frames, proper sampling and filtering must be applied when fetching values from those previous frame texture stacks to avoid any aliasing effect, as well as to account for discontinuities between layers that might lead to inconsistencies in the simulation. When fetching values in the previous frame stack, all layers are sampled and depth values in the stack are used to prune out invalid samples.

Bleeding Rendering: During the final rendering, blood information is fetched from the simulation results using a custom texture filtering approach similar to the one used in inter-frame propagation, since the changing observer viewpoint does not in general match the fluid simulation viewpoint and texture resolution. At that stage, we first displace the vertices on the GPU to account for the amount of blood at that location. This is mostly a simple step, but in areas at risk of numerical uncertainty errors, such as areas where the mesh is almost aligned with gravity (e.g. the left side of the surface in Fig. 3), a validation test is applied to insure that the displaced mesh is coherent with the elevation computed by the image-based simulation. This allows for example multiple vertices located on a vertical wall to be displaced to the same position when under the fluid surface. For meshes that are too coarse to support this approach, we have experimented with surface subdivision in a GPU geometry shader based on PN triangles [14]. This allowed the mesh displacement to be sufficiently fine, and reduced performance only slightly more than if using the finer mesh without GPU subdivision. To deal with lighting in areas where the blood has accumulated, we compute on the GPU a set of n surface normal maps (one per front-facing layer) by taking the derivative on the n layers of depth map, again taking into account the surface connectivity between the maps. These normals progressively replace the ones computed from the mesh as a factor of blood accumulation. Such normal maps are displayed in Figure 4. In the same manner, we blend the original tissue color with the blood color as a function of the amount of fluid. The blending parameters are variables in the simulator.

3 Results and Discussion

For the results discussed here (Fig. 1), we used a stack of five depth layers, three front facing and two back facing, and a monoscopic final rendering window at 1280×1024 resolution. Tests were run on a quad-core Intel i7 PC and with

one core of a NVIDIA GeForce GTX 295 GPU. Figure 1 illustrates the typical steps followed by a trainee learning to perform cauterization and aspiration for hemostasis during tumor resection. Videos are available at http://Neurotouch.ca. Since the mesh may need to be rendered with the fluid many times per frame due to special effects such as cast shadows, times required for fluid propagation and rendering are given separately. For the 128×128 simulation grid used in the videos, the fluid simulation time is of 3.2 ms per frame + 1.7 ms for rendering of the global scene including the vertex displacement of a 30K polygon deformable surface mesh. With much larger 512×512 blood and depth textures, blood-related simulation time only increases to $9 + 4.4 = 13.4$ ms. As can be seen from those numbers, the algorithm allows interactive simulation of the bleeding process at a relatively large simulation resolution. But for our application we get a bleeding that is more realistic by using smaller textures and better texture sampling/filtering at the final stage of rendering. This allows blood to propagate more quickly and at a more realistic granularity given the surgery scale (3-5 cm). Alternately, one could run two iterations of fluid simulation for each iteration of rendering. The complete visual simulation including bleeding, shadows and depth of field effects runs at a refresh rate of 60 Hz with more than 15 rendering passes per frame.

The current approach does have some limitations. Performance will diminish rapidly as the number of required layers increases, so the technique may not be practical for very intricate geometric models. This is due mainly to the fact that the custom texture filtering associated with the rendering and inter-frame texture fetches require comparison and combination of all layers in the stack. There would also be a benefit in expanding the method to be able to change the viewpoint for the depth peeling, to increase the rendering quality when observer viewpoints are too far away from the gravity axis. Finally, the vertex displacement approach used for the final rendering can lead to discontinuities in the geometric mesh in specific cases. These issues do not affect the validity of the simulation, but can lead to visual artefacts that need to be mitigated by modifying color and normal attributes on the mesh. This will happen when an enclosed pool of fluid gets filled and a part of the mesh closes on itself due to the blood displacement (e.g. Fig. 3(b)). A vertex displacement strategy also makes it difficult to assign different levels of transparency to the fluid columns in the final rendering, which may be an issue for rinsing/dilution simulation.

4 Conclusion

We have presented a method to efficiently integrate grid-based surface fluid simulation to context of neuro-surgery simulation. The proposed technique allows the realistic simulation of surface bleeding, including aspiration and cauterization, over the entire operating field at a low computational cost. Future work will include the implementation of a ray-traced final render pass to allow a more complete set of visual features for the bleeding simulation, thus enabling the simulation of other important aspects of surgery. Jet bleeding and fluid falling

from overhangs could be added using localized special effects or small amounts of particles at relatively low cost. Finally, the method proposed here could be adapted to other contexts, from gaming to scientific simulation.

References

1. Andersson, L.: Real-Time Fluid Dynamics for Virtual Surgery. Master's thesis, Engineering Physics Program, Chalmers University of Technology (2005)
2. Basdogan, C., Ho, C.H., Srinivasan, M.A.: Simulation of tissue cutting and bleeding for laparoscopic surgery using auxiliary surfaces. In: Medicine Meets Virtual Reality (MMVR7) Conference, pp. 38–44. IOS Press, Amsterdam (1999)
3. Everitt, C.: Interactive order-independent transparency. White paper, NVIDIA (2001)
4. Kass, M., Miller, G.: Rapid, stable fluid dynamics for computer graphics. In: 17th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 1990, pp. 49–57. ACM, New York (1990)
5. Kerwin, T., Shen, H.W., Stredney, D.: Enhancing realism of wet surfaces in temporal bone surgical simulation. IEEE Transactions on Visualization and Computer Graphics 15(5), 747–758 (2009)
6. Mei, X., Decaudin, P., Hu, B.G.: Fast hydraulic erosion simulation and visualization on GPU. In: 15th Pacific Conference on Computer Graphics and Applications (PG 2007), pp. 47–56. IEEE Computer Society, Los Alamitos (2007)
7. Müller, M., Schirm, S., Teschner, M.: Interactive blood simulation for virtual surgery based on smoothed particle hydrodynamics. Technol. Health Care 12, 25–31 (2004)
8. Myers, K.: HLSL blood shader gravity maps. White paper, NVIDIA (July 2004)
9. O'Brien, J., Hodgins, J.: Dynamic simulation of splashing fluids. In: Computer Animation 1995, pp. 198–205. IEEE Computer Society, Washington (1995)
10. Pang, W.M., Qin, J., Chui, Y.P., Wong, T.T., Leung, K.S., Heng, P.A.: Orthopedics surgery trainer with PPU-accelerated blood and tissue simulation. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part II. LNCS, vol. 4792, pp. 842–849. Springer, Heidelberg (2007)
11. Qin, J., Chui, Y.P., Pang, W.M., Choi, K.S., Heng, P.A.: Learning blood management in orthopedic surgery through gameplay. IEEE Computer Graphics and Applications 30(2), 45–57 (2010)
12. Raghupathi, L., Devarajan, V., Eberhart, R., Jones, D.: Simulation of bleeding during laparoscopic herniorrhaphy. In: Medicine Meets Virtual Reality (MMVR12) Conference, pp. 382–387. IOS Press, Amsterdam (2002)
13. Rianto, S., Li, L., Hartley, B.: Fluid dynamic simulation for cutting in virtual environment. In: Skala, V. (ed.) 16th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision. University of West Bohemia (2008)
14. Vlachos, A., Peters, J., Boyd, C., Mitchell, J.L.: Curved PN triangles. In: 2001 Symposium on Interactive 3D Graphics, I3D 2001, pp. 159–166. ACM, New York (2001)
15. Šťava, O., Beneš, B., Brisbin, M., Křivánek, J.: Interactive terrain modeling using hydraulic erosion. In: Gross, M., James, D. (eds.) 2008 Eurographics/ACM SIGGRAPH Symposium on Computer Animation, SCA 2008, pp. 201–210. Eurographics Association, Aire-la-Ville (2008)

An Application-Dependent Framework for the Recognition of High-Level Surgical Tasks in the OR

Florent Lalys[1,2,3], Laurent Riffaud[1,2,3,4], David Bouget[1,2,3], and Pierre Jannin[1,2,3]

[1] INSERM, U746, Faculté de Médecine CS 34317,
F-35043 Rennes Cedex, France
[2] INRIA, VisAGeS Unité/Projet, F-35042 Rennes, France
[3] University of Rennes I, CNRS, UMR 6074, IRISA, F-35042 Rennes, France
[4] Department of Neurosurgery, Pontchaillou University Hospital,
F-35043 Rennes, France

Abstract. Surgical process analysis and modeling is a recent and important topic aiming at introducing a new generation of computer-assisted surgical systems. Among all of the techniques already in use for extracting data from the Operating Room, the use of image videos allows automating the surgeons' assistance without altering the surgical routine. We proposed in this paper an application-dependent framework able to automatically extract the phases of the surgery only by using microscope videos as input data and that can be adaptable to different surgical specialties. First, four distinct types of classifiers based on image processing were implemented to extract visual cues from video frames. Each of these classifiers was related to one kind of visual cue: visual cues recognizable through color were detected with a color histogram approach, for shape-oriented visual cues we trained a Haar classifier, for texture-oriented visual cues we used a bag-of-word approach with SIFT descriptors, and for all other visual cues we used a classical image classification approach including a feature extraction, selection, and a supervised classification. The extraction of this semantic vector for each video frame then permitted to classify time series using either Hidden Markov Model or Dynamic Time Warping algorithms. The framework was validated on cataract surgeries, obtaining accuracies of 95%.

Keywords: Surgical phase, digital microscope, cataract surgeries, DTW.

1 Introduction

The field of surgical process analysis and modelling has recently gained much interest. Due to the technologically rich environment of the Operating Room (OR), a new generation of computer-assisted surgical (CAS) systems has appeared. As a result of these new systems, a better management, safety, and comprehension of the surgical process is needed. For such purposes, systems should rely on a context-aware tool, which knows the score to be played for adapting assistance accordingly. The challenge is therefore to assist surgery through the understanding of OR activities, which could be introduced in CAS systems. Clinical applications also include evaluation and training of surgeons, the creation of context-sensitive user interfaces, or the generation of automatic post operative reports.

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 331–338, 2011.

The goal is to collect signals from the OR and automatically derive a model. While it is possible to design such model manually, there are advantages of automating this process, mainly because manual work is time-consuming and can be affected by human bias and subjectivity. Due to the increasing number of sensors in the OR, the automatic extraction of data is now easier. Based on these signals, it's possible to recognize high-level tasks and hence avoid any additional installation of materials. Among all sensors, teams recently focused on videos coming from cameras already used in the clinical routine, which are a rich source of information. Compared to other data extraction techniques, it uses a source that does not have to be controlled by humans, automating the surgeons' assistance without altering the surgical routine.

Current work has made progress in classifying and automating the recognition of high-level tasks in the OR based on videos. Using laparoscopic videos, Speidel et al. [1] focused on surgical assistance by identifying 2 scenarios: one for recognizing risk situations and one for selecting adequate images for visualization. Their analysis was based on augmented reality and computer vision techniques. Lo et al. [2] used vision to segment the surgical episode. They used color segmentation, shape-from-shading techniques, and optical flows for tracking instruments. These features, combined with other low-level cues, were integrated into a Bayesian framework. Klank et al. [3] extracted image features for scene analysis and frame classification. A crossover combination was used for selecting features, while Support Vector Machines (SVMs) were used for the classification. Blum et al. [4] automatically segmented the surgery into phases. A Canonical Correlation Analysis was applied based on tool usage to reduce the feature space, and the modeling of resulting feature vectors was performed using Dynamic Time Warping (DTW) and Hidden Markov Model (HMM). Bhatia et al. [5] analyses overall OR view videos. After identifying 4 states of a common surgery, relevant image features were extracted and HMMs were trained to detect OR occupancy. Padoy et al. [6] also used external OR videos to extract low-level image features through 3D motion flows combined with hierarchical HMMs to recognize on-line surgical phases. In robotic using the Da Vinci, Voros and Hager [7] used kinematic and visual features to classify tool/tissue interactions. Similarly, Reiley and Hager [8] focused on the detection of subtasks for surgical skill assessment.

In a previous work [9], using neurosurgical videos, we proposed to extract surgical phases by combining a feature extraction process with HMM. In this paper, we extend this approach by proposing an application-dependent framework that can be adaptable to any type of surgeries. The idea is first to extract visual cues that can be helpful for discriminating high-level tasks. The visual cues are detected by specific image-based classifiers, obtaining a semantic signature for each frame. Then, these time series are aligned with a reference surgery using DTW algorithm to recognize surgical phases. Compare to traditional video understanding algorithms, this framework extracts application-dependant visual cues that are generic. The combination of image-based analysis and time series classification allows getting high recognition rates. We evaluated our framework with a dataset of cataract surgeries through cross-validation studies, and compared results of the DTW approach with the HMM classification.

2 Materials and Methods

2.1 Application-Dependant Visual Cues

Four classifiers based on different image processing tools were implemented (Fig. 1.). Each of these classifiers was related to one kind of possible visual cue. Visual cues recognizable through color were detected using a color histogram approach. For each shape-oriented visual cue such as the recognition of a specific object, a Haar classifier was trained. For texture-oriented visual cues, we used a bag-of-word approach using Scale Invariant Feature Transform (SIFT) descriptors, and finally for other visual cues that don't match these descriptions, we used an image classification approach including a feature extraction, selection and a classification with SVM.

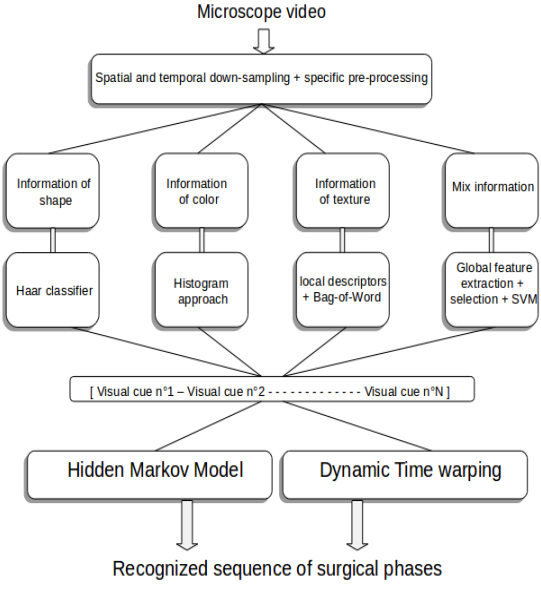


Fig. 1. Framework of the recognition system

Color-oriented visual cues: The color is one of the primary features used to represent and compare visual content. Especially, color histograms have a long history as a method for image description, and can also be used for identifying color shade through images. Here we used the principle of histogram classification to extract color-oriented visual cues, by creating a training image database composed of positive and negative images. Two complementary color spaces were extracted: RGB and HSV space. For quantifying similarities between histograms, we used the correlation.

Shape-oriented visual cues: We used here a Viola-Jones object detection framework [10], mainly used to detect specific object within images. The basic idea is to create a classifier based on features selected by AdaBoost. Weak learners of the algorithm are based on the Haar-like rectangular features, comparing the sum of intensities in adjacent regions inside a detection window. Then, strong learners are arranged into a classifier cascade tree in complexity order. The cascade classifier is therefore composed of stages, each one containing a strong learner. During the detection phase, a window looks through the image with different scales and positions. The idea is to determine at each stage if a given sub-window may be the searched object or not. The false positive rate and the detection rate are thus the product of each rate at each stage.

Texture-oriented visual cues: For whole-image categorization tasks, bag-of-visual-words (BVW) representations, which represent an image as an orderless collection of

local features, have demonstrated impressive performances. The idea of BVW is to treat images as loose collections of independent patches, sampling a representative set of patches from the image, evaluating a descriptor vector for each patch independently, and using the resulting distribution of samples in descriptor space as a characterization of the image. A bag of keypoints is then expressed as a histogram recounting the number of occurrences of each pattern in a given image. For the texture analysis, we used the SIFT [11] descriptors.

Other visual cues: We already presented this approach in a previous paper [12]. Each frame was represented by a signature composed of low-level spatial features (RGB space, co-occurrence matrix with Haralick descriptors [13], spatial moments [14], and Discrete Cosine Transform (DCT) [15] coefficients). This signature was then reduced by feature selection. For that purpose, we fused a filter and a wrapper approach by using the union of both selection results. The RFE-SVM [16] and the mutual information (MI) [17] were chosen for the wrapper and the filter method respectively, keeping the 40 first features. Finally, a SVM was applied to extract the binary cue.

2.2 Time Series Classification

A binary semantic signature was extracted from each frame, composed of the recognized visual cues. We used the DTW algorithm [18] to classify these time series in a supervised way. The objective of DTW is to compare two sequences by computing an optimal match. These sequences may be time-series composed of feature sequences sampled at equidistant points in time. A local cost measure is needed to compare features. We used here the Hamming distance. To compare each surgery, we created an average surgery with the method described in [19]. Every query surgery was first processed to extract visual cues, and then the time series were compared to the average one. Once warped, the phases of the average surgery are transposed to the query one. We also used the Itakura parallelogram global constraint that limits the warping path to be within a parallelogram.

2.3 Data-Set

Our framework was evaluated on cataract surgeries. 20 cataract surgeries from the Hospital of Munich were included to the study (mean surgical time: 15 min). Videos were recorded using the OPMI Lumera surgical microscope (Carl Zeiss) with a resolution of 720 x 576 at 25 fps. We downsampled the videos to 1 fps, and spatially downsampled by a factor 8 with a 5-by-5 Gaussian kernel. Eight surgical phases were defined (Fig. 2). Additionally, five binary visual cues were chosen: the pupil color range (orange or black), the presence of antiseptic, of the knife, of the IOL instrument, and the global aspect of the cataract. Combinations of these five binary cues are informative enough to discriminate all 8 phases. The pupil color range and the presence of the antiseptic were extracted using color histograms. The knife was recognized using a Haar classifier. The IOL instrument was not identifiable through only color or shape analysis, that's why we chose the fourth approach using global

spatial feature extraction and SVM classification. Finally, the global aspect of the cataract was recognized using the BVW approach. For this detection as well as for the pupil color range classification, a step of pupil segmentation was first applied, using preprocessing steps composed of dilation/erosion operations and a Hough transform.

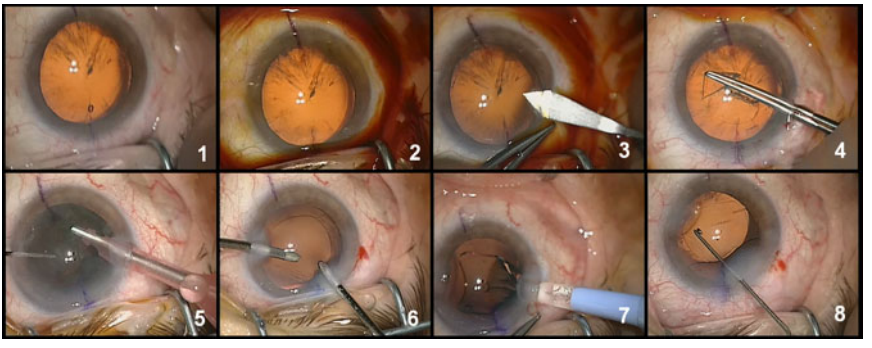


Fig. 2. Example of typical digital microscope images for the eight phases: 1-preparation, 2-betadine injection, 3-corneal incision, 4-capsulorhexis, 5-phacoemulsification, 6-cortical aspiration of the remanescent lens, 7-implantation of artificial IOL, 8-adjustment of the IOL

2.4 Cross-Validation

The initial work of phase and visual cue labeling was performed for each video. From each video, we randomly extracted 100 frames, getting an image database composed of 2000 labeled images. We then evaluated both aspects of our framework. First, every visual cue detection was assessed trough 10-fold cross-validation studies, by dividing the image database into 10 random subsets. Then, we evaluated the global framework with the same procedure. At each stage, 18 videos (and their corresponding frames from the image database) were used for training and recognitions were made on the 2 others. For this validation, we computed the Frequency Recognition Rate (FRR). We also validated the added-value of the DTW algorithm, by comparing with an HMM approach, described in previous studies [9].

3 Results

Results of the cross-validation studies (Tab. 1.) showed that very good accuracies were obtained for visual cues with quite low standard deviations. The best recognition was obtained for the detection of the antiseptic, with a recognition rate of 98.5%, whereas the lower rate was obtained for the IOL instrument recognition (94.8%). An example of a DTW computation is shown on Fig. 3. Small errors occur in the phase transitions, but the global FRR stay high (~93%). Tab. 2. shows the global accuracy of the framework, using DTW or HMM approach. The global validation study with DTW showed a **mean FRR of 94.8%**, with a **min of 90.5%** and a **max of 98.6%.**

Table 1. Mean accuracy and standard deviation (Std) for the recognition of the 5 binary visual cues, computed on the entire video dataset

	Pupil color range	Presence antiseptic	Presence Knife	Presence IOL instrument	Cataract aspect
Accuracy (%)	96,5	98,5	96,7	94,8	95,2
Std (%)	3,7	0,9	3,4	1,1	1,8

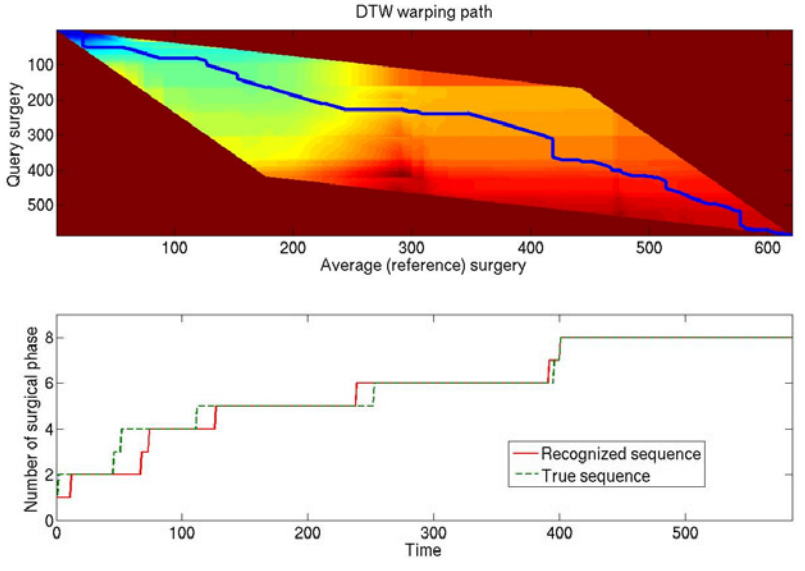


Fig. 3. Distance map of two surgeries and dedicated warping path using the Itakura constraint (up), along with the transposition of the surgical phases (down)

Table 2. Mean, minimum and maximum FRR of the HMM and DTW studies

	FRR (Std)	Minimum (%)	Maximum (%)
HMM (%)	92,2 (6,1)	84,5	99,8
DTW (%)	94,8 (3,7)	90,5	98,6

4 Discussion

4.1 Visual Cues and DTW

Combining with state-of-the-art techniques of visual cues recognition, DTW showed very good performance and allows further promising works on high-level tasks recognition in surgery. The comportment in color, texture and shape of the visual cues are intuitively known, allowing the classifiers to be effective. This approach turns out to be as generic as possible, and adaptable to any type of surgery. However, one limitation of the DTW algorithm is that it can't be used on-line, because the entire surgery is needed in order to find the optimal path. However, first results showed that the DTW algorithm was quite better for classifying times series data than the HMM.

4.2 Microscope Video Data

The real added value of the project lies in the use of microscope videos. This device is not only already installed in the OR, but it has also not to be monitored by the staff. Compared to other additional sensors, this allows the recognition to be fully automatic and independent. Moreover, microscope video data are reproducible within a same surgical environment and image features are invariant to task distortion [20]. Due to facilities differences between surgical departments, the system could not be flexible. The solution would be to train dedicated databases for each department, which would be adapted to the corresponding surgical environment and microscope scene layout.

4.3 Clinical Applications

The automatic recognition of surgical phases might be helpful for various applications. Purposes are generally to bring an added value to the surgery or to the OR management. This work could be integrated in an architecture that would take in real-time the microscope videos as input and transform it into information helping the decision making process, and driving context-sensitive user interfaces. Off-line, surgical videos would be very useful for learning and teaching purposes given their automatic indexation. Moreover, we could imagine the creation of pre-filled post operative reports that will have to be completed by surgeons. The recognition of lower level information, such as gestures, is difficult with microscope videos only. In future works, lower-level information such as surgeon's gestures will have to be detected to create multi-layer architectures.

5 Conclusion

We proposed in this paper a framework that automatically recognizes surgical phases from microscope videos. The first step of the framework is the definition of several visual cues for extracting semantic information and therefore characterizing every frame. Then, time series models allow an efficient representation of the problem by modeling time varying data. This association permits to combine the advantages of all methods for better modeling. We tested the framework on cataract surgeries, where 8 phases and 5 visual cues were defined by an expert, getting global accuracies of 95%. This recognition process is a first step in the construction of context-aware surgical systems, opening perspectives for a new generation of CAS systems.

Acknowledgements. The authors would like to acknowledge the financial support of Carl Zeiss Meditec.

References

1. Speidel, S., Sudra, G., Senemaud, J., Drentschew, M., Müller-stich, B., Gun, C., Dillmann, R.: Situation modeling and situation recognition for a context-aware augmented reality system. In: Progression in Biomed. Optics and Imaging, vol. 9(1), p. 35 (2008)
2. Lo, B., Darzi, A., Yang, G.: Episode Classification for the Analysis of Tissue/Instrument Interaction with Multiple Visual Cues. In: Ellis, R.E., Peters, T.M. (eds.) MICCAI 2003. LNCS, vol. 2878, pp. 230–237. Springer, Heidelberg (2003)

3. Klank, U., Padoy, N., Feussner, H., Navab, N.: Automatic feature generation in endoscopic images. Int. J. Comput. Assist. Radiol. Surg. 3(3-4), 331–339 (2008)
4. Blum, T., Feußner, H., Navab, N.: Modeling and Segmentation of Surgical Workflow from Laparoscopic Video. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A., et al. (eds.) MICCAI 2010. LNCS, vol. 6363, pp. 400–407. Springer, Heidelberg (2010)
5. Bhatia, B., Oates, T., Xiao, Y., Hu, P.: Real-time identification of operating room state from video. In: AAAI, pp. 1761–1766 (2007)
6. Padoy, N., Blum, T., Feuner, H., Berger, M.O., Navab, N.: On-line recognition of surgical activity for monitoring in the operating room. In: Proc's of the 20th Conference on Innovative Applications of Artificial Intelligence (2008)
7. Voros, S., Hager, G.: Towards "real-time" tool-tissue interaction detection in robotically assisted laparoscopy. Biomed. Robotics and Biomechatronics, 562–567 (2008)
8. Reiley, C., Hager, G.: Decomposition of robotic surgical tasks: an analysis of subtasks and their correlation to skill. In: M2CAI Workshop, MICCAI 2009 (2009)
9. Lalys, F., Riffaud, L., Morandi, X., Jannin, P.: Surgical phases detection from microscope videos by combining SVM and HMM. In: Menze, B., Langs, G., Tu, Z., Criminisi, A. (eds.) MICCAI 2010. LNCS, vol. 6533, pp. 54–62. Springer, Heidelberg (2011)
10. Viola, P. and Jones, M.: Rapid real-time face detection. IJCV, 137-154 (2004)
11. Lowe, D.G.: Object recognition from scale-invariant features. In: ICCV 1999, vol. 2, pp. 1150–1157 (1999)
12. Lalys, F., Riffaud, L., Morandi, X., Jannin, P.: Automatic phases recognition in pituitary surgeries by microscope images classification. In: Navab, N., Jannin, P. (eds.) IPCAI 2010. LNCS, vol. 6135, pp. 34–44. Springer, Heidelberg (2010)
13. Haralick, R., Shanmugam, K., Dinstein, I.: Textural features for image classification. IEEE Trans. on Systems, Man, and Cybernetics 3(6), 61–621 (1973)
14. Hu, M.: Visual pattern recognition by moment invariants. Trans. Inf. Theory 8(2), 79–87 (1962)
15. Ahmed, N., Natarajan, T., Rao, R.: Discrete Cosine Transform. IEEE Trans. Comp., 90–93 (1974)
16. Guyon, I., Weston, J., Barhill, S., Vapnik, V.: Gene selection for cancer classification using support vector machine. Machine Learning 46, 389–422 (2002)
17. Hamming, R.W.: Coding and Information Theory. Prentice-Hall Inc., Englewood Cliffs (1980)
18. Sakoe, H., Chiba, S.: Dynamic programming algorithm optimization for spoken work recognition. In: IEEE Trans. Acoustic, Speech and Signal Proc., vol. 26(1), pp. 43–49 (1978)
19. Ahmadi, S.-A., Sielhorst, T., Stauder, R., Horn, M., Feussner, H., Navab, N.: Recovery of surgical workflow without explicit models. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4190, pp. 420–428. Springer, Heidelberg (2006)
20. Bouarfa, L., Jonker, P., Dankelman, J.: Discovery of high-level tasks in the operating room. J. Biomedical Informatics 44(3), 455–462 (2010)

Biomechanical Simulation of Electrode Migration for Deep Brain Stimulation

Alexandre Bilger, Jérémie Dequidt, Christian Duriez, and Stéphane Cotin

SHAMAN Group, INRIA Lille North Europe

Abstract. Deep Brain Stimulation is a modern surgical technique for treating patients who suffer from affective or motion disorders such as Parkinson's disease. The efficiency of the procedure relies heavily on the accuracy of the placement of a micro-electrode which sends electrical pulses to a specific part of the brain that controls motion and affective symptoms. However, targeting this small anatomical structure is rendered difficult due to a series of brain shifts that take place during and after the procedure. This paper introduces a biomechanical simulation of the intra and postoperative stages of the procedure in order to determine lead deformation and electrode migration due to brain shift. To achieve this goal, we propose a global approach, which accounts for brain deformation but also for the numerous interactions that take place during the procedure (contacts between the brain and the inner part of the skull and falx cerebri, effect of the cerebro-spinal fluid, and biomechanical interactions between the brain and the electrodes and cannula used during the procedure). Preliminary results show a good correlation between our simulations and various results reported in the literature.

1 Introduction

Deep Brain Stimulation (DBS) is a modern surgical treatment of brain disorders such as Parkinson's disease or dystonia. This procedure consists in the placement of a micro-electrode in the subthalamic area, deep into the brain. The placement of the electrode is crucial to maximize outcomes and to prevent adverse effects. This placement is achieved in two main stages: first, pre-operative medical images of the patient are combined with the use of a stereotactic frame (and sometimes an atlas of the brain) to determine the target coordinates and optimal trajectory for the electrode(s). Second, the patient is taken to the operating room where a macro-electrode attached to a thin wire is inserted into the brain according to the planned trajectory. However, a combination of brain shift and a certain inaccuracy in the exact location of the target area require to test the area next to the planned target to optimize the placement of the electrode. This testing is performed by recording the brain activity using the macro-electrode. This entire process takes several hours, and once the appropriate area has been identified the final (micro) electrode will be left in place and secured to the skull. However, two major problems arise during this process: first, the planning stage does not account for the brain shift that takes place during surgery. Depending on

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 339–346, 2011.

the amplitude of the brain shift, the effective location for the electrode can be quite remote from the planned location (5 mm or more) [13] requiring extensive mapping of the area to determine the final location. As a consequence, this stage of the procedure usually lasts several hours, while the patient is awake, and increases the likelihood of complications due to the number of insertions to reach the targeted area [4]. The second problem, also linked to the amplitude of the brain shift, takes places several days or weeks after the surgery. As reported in [13] a post-operative electrode displacement and deformation may appear as the brain returns to its initial position when the subdural air introduced during surgery has resolved (see Fig.1). This hinders the efficiency of the procedure because upward migration of the electrode may fail to correctly stimulate the subthalamic area.

In this context, our objective is to propose a global approach that can model these two phenomena in order to adjust the planned trajectory, determine a potential post-operative electrode migration, and propose alternative strategies to minimize its amplitude. As a first step in this direction, this paper introduces an original and unified approach to model the brain behavior during a DBS procedure. The focus of the paper is not set on a specific biomechanical model of the brain but rather on a complete framework that is able to simulate intra-cranial fluid loss, subdural air invasion, brain shift and electrode migration and curvature. This work includes the following contributions: mechanical models of the brain and the devices (cannula and electrode); mechanical interactions between the brain and these devices (in particular electrode deformation when the brain reverts to its initial shape); influence of the surrounding cerebro-spinal fluid (CSF) and air invasion in the skull on the brain shift (including the asymmetry of the brain shift). Results of the simulation exhibit qualitatively consistent results compared to these various points reported in the literature.

This paper is organized as follows: section 2 reviews previous works related to simulation of brain and DBS, section 3 details the various elements of our approach, and finally section 4 provides preliminary results.

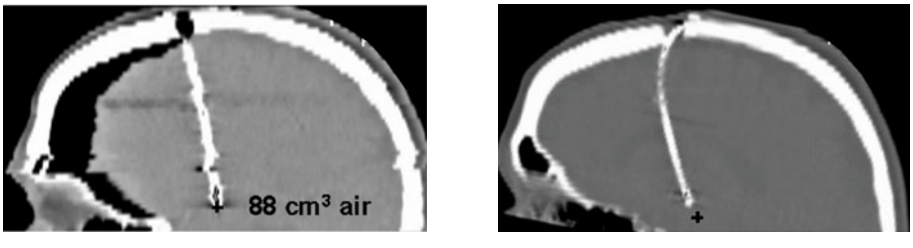


Fig. 1. Post-operative (left) and follow-up (right) CT scans. The post-operative scan illustrates the brain shift at the end of the procedure. The follow-up scan emphasizes the deformation of the electrode due to the inverse brain shift, leading to an upward migration of the electrode away from its initial location after the craniotomy (black cross). For large brain shifts, the electrode can move of up to 5 mm. Courtesy of [13].

2 Previous Works

As stated in the introduction, the craniotomy-induced brain shift is one of the main factors that impact the planning, duration and complexity of DBS procedures. Brain shift also impacts many other neurosurgical procedures. In the literature related to brain shift, the focus is on measuring or computing the displacement field of the brain in order to accurately track the targeted area (for instance a tumor). Most of these approaches rely on more or less advanced biomechanical models of the brain, usually guided by coarse intra-operative information such as ultrasonography [2], intra-operative MRI [3] or laser-range scanner [15]. However, such approaches, that typically use per-operative information to post-correct the deformation of the brain to match measured surface displacement rely on oversimplified boundary conditions. These approaches fail to correctly reproduce the complex interactions that take place inside the skull. While this is acceptable when only local tissue deformation is sought (e.g. tumor removal near the brain surface), this becomes a limitation for a more global estimation of the deformation (as needed for DBS). They also do not consider post-operative recovery of the brain which plays a crucial role in deep brain stimulations because of possible electrode migration, which has been recently highlighted by [6,13].

Estimating the final position (e.g. after brain recovery) of the electrode requires to take into account various elements in the simulation: a) biomechanical brain model, b) bilateral and unilateral boundary conditions with the skull and falx cerebri, c) interactions between the brain, the cannula and the electrode(and its wire) and finally d) loss of CSF due to the craniotomy. To our knowledge, such a complete and unified framework has not been addressed in the literature. Some works have proposed advanced biomechanical brain models based on linear elasticity [3] or non-linear visco-elasticity [11]. The influence of boundary conditions is emphasized in [14] where three scenarios for brain-skull boundary conditions are compared and experiments show that best results are achieved when brain motion is allowed in the cranial cavity. Regarding the insertion of the cannula or electrode in the brain, the closest works are in the field of brachytherapy (the reader may refer to [1] for a survey on the insertion of needles into soft tissues). Finally regarding the influence of CSF, Lunn *et al* [8] propose to model the brain as a porous media and use consolidation theory to take into account the CSF in the brain.

3 Simulation Framework and Methods

The principal contribution of this paper is to propose a global, physics-based approach that models the main (bio)mechanical phenomena that can be observed during and after a DBS procedure, i.e.:

- **Asymmetric brain shift** which can be observed during the first electrode implantation. It is due to a unilateral air invasion on the side of the first craniotomy, resulting in a contralateral brain shift.

– **Two-stage brain shift** which appears after removing the cannula. While most of the brain shift takes place right after the craniotomy and before the insertion of the cannula and macro-electrode, the air invasion continues to occur during the lengthy testing process. However, the cannula being rigid, it locally reduces brain shift. When the cannula is removed, an additional shift of the brain can sometimes be observed and can drive to a first deformation of the flexible electrode.
– **Inverse brain shift** which takes place several days or weeks after the procedure. As the air invasion is resolved and new CSF has been produced, the brain recovers its initial shape and position. During this *inverse brain shift*, a deformation is also applied to the micro-electrode and its wire. This interaction creates a relative motion between the electrode and the brain, along the axis of the electrode. If this motion is important, the electrodes are no longer able to stimulate the correct subthalamic area.

3.1 Cerebro-Spinal Fluid Model

Brain shift takes place due to a loss of cerebro-spinal fluid and subdural air invasion after the skull opening(see Fig.3). Before the craniotomy, the buoyancy force created by the CSF balances the gravity force acting on the brain. But after a loss of fluid through the burr hole, a part of the cerebral tissue is left above the remaining CSF surface and undergoes only gravity. This unbalance can deform and move the brain significantly. Reported measures indicate that even deep brain structures (e.g. the targets of a DBS such as the Subthalamic Nucleus) may shift of up to 5mm [12].

The external force created on the surface of the brain by the CSF is computed as $\mathbf{f}_{CSF} = \iint_S \rho g h(P)\,\mathrm{d}\mathbf{S}$ with ρ is the density of CSF, g the norm of the gravity and h the distance between a point P on the brain surface and the fluid level. This force is computed on each triangle S of the brain mesh (see below) that corresponds to the immersed surface. Results are illustrated in section 4.

3.2 Constrained Deformations of the Brain

The anatomy of the brain and skull is based on a generic atlas and is not issued from patient data. The deformations of the brain are modeled using the finite element method. The volume of the brain hemispheres is meshed as a set of tetrahedral elements, and different constitutive models can be used to describe the brain deformation. In the current results, we have used both a co-rotational approach (to capture the geometrical non-linearities of the deformations) and a Saint Venant-Kirchhoff model to describe a hyperelastic material. While there are other constitutive laws that have been proposed for modeling brain tissue, we want to emphasize in this section the importance of correctly accounting for the complex boundary conditions, which can have a larger impact on the final deformation than the choice of a particular deformation model. In the particular case of the CSF influence on brain deformation, the change of forces acting on the brain leads to a brain shift, while the falx cerebri which separates the two

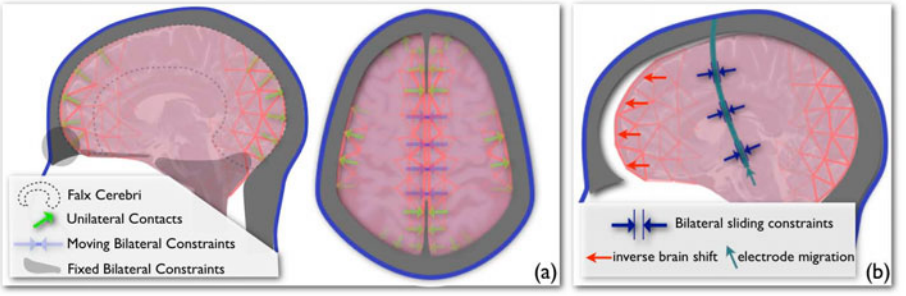


Fig. 2. Illustration of the different models used by the simulation and their relative boundary conditions. The use of complex constraints, combined with a FEM modeling technique permits to capture some of the key characteristics of a DBS procedure.

brain hemispheres limits the influence of the CSF loss and air invasion to the side of the craniotomy. This relatively complex process is simulated by modeling independently each hemisphere of the brain and by precisely defining boundary conditions between the brain hemispheres themselves, and between the brain and skull or falx cerebri.

Independently of the choice of the deformation model, we end up with the following differential system of non-linear equations $\mathbf{Ma} = \mathbf{f}(\mathbf{x}, \mathbf{v}) + \mathbf{p} + \mathbf{f}_{CSF} + \mathbf{H}^T\lambda$ where $\mathbf{M}$ is the mass matrix, $\mathbf{f}$ gathers the internal forces. $\mathbf{a}$, $\mathbf{v}$ and $\mathbf{x}$ are respectively the acceleration, the velocity and the position of the nodes from the mesh. The forces $\mathbf{p}$ are exerted by the gravity and $\mathbf{f}_{CSF}$ by the CSF. Finally, $\mathbf{H}^T\lambda$ gathers constraints response resulting from unilateral contacts and bilateral constraints (e.g. see Fig. 2 for brain-falx cerebri and brain-skull contacts; moving bilateral constraints between the two brain hemispheres; fixed bilateral constraints near the area of the optic nerves and the brainsterm). An implicit integration scheme (Backward Euler) is used to enforce stability even when using large time steps.

3.3 Brain-Electrode Interactions

With the objective to estimate the relative motion between the electrode and the brain during the *inverse brain shift*, and define new strategies to compensate for that, it is essential to accurately model the mechanical interactions between the cannula, the electrode and the brain tissue during the different stages of the insertion and deformation. The electrode and cannula are modeled using serially-linked beams elements as done in [5] for coil modeling. This non-linear elastic model can be parametrized to reproduce the resistance to bending, stretching and torsion of the two types of devices. The mechanical coupling between the devices and brain is controlled using constraints that are solved using additional Lagrange Multipliers, following the approach presented in [5]. A set of sliding point constraints is positioned in the brain model along the path of the cannula; and a similar set of constraints is used for the electrode. For solving the coupling

Fig. 3. Asymmetrical brain shift simulation: the brain shift is more important on the side where the craniotomy first takes place, as illustrated in the left most images (courtesy of [9]). The rightmost images show the evolution of brain shift during our simulation: (a) no brain shift; (b) after a right craniotomy; and (c) after both craniotomies.

between the models, Lagrange Multipliers are set for both the brain tissue model and beam models. The final values of the Lagrange Multipliers are solved using a Mixed-Complementarity Problem solver (MCP).

4 Results

In this section we present a series of results for the main steps of the procedure described in section 3. The conditions and parameters used in the simulation reproduce as closely as possible the conditions and parameters reported in the literature to facilitate comparisons, i.e : the patient lies in the supine position with the head elevated at 10 to 20 degrees; the craniotomy is performed first on the right side of the skull; a first brain shift takes place before the cannula is inserted through the right hemisphere; the electrode and wire are then inserted through the cannula; the cannula is removed and the electrode remains in the brain but is slightly shifted due to secondary brain shift; the CSF is restored and the air is removed to simulate post-operative conditions several weeks after the surgery. The same process is repeated for the left side of the brain. As values reported in the literature for the Young's modulus range from 2,100 [10] to 40,000 Pa[7], we use, for both the co-rotational and Saint Venant Kirchhoff models a Young's modulus $E = 6000Pa$ and a Poisson ratio $\nu = 0.45$ for a total mass of the brain of $1.4kg$. The density of CSF ρ is set to 1000 kg/m^3. Figure 3 illustrates the simulation of the asymmetric brain shift. During the surgery and even after the loss of CSF, the brain shifts continuously because of the pressure due to air invasion. However, the rigid material of the cannula prevents the brain motion. That is why, the removal of the cannula causes a second but minor brain shift. This effect involves an anteroposterior deformation of the implanted wire, as depicted in figure 4. Several days after surgery no more air is in the cranial space and the CSF has been restored. We simulate this effect by modifying the CSF level. The resulting deformations of the brain are computed using our FEM approach. As the electrode and its wire are constrained within the brain (only sliding is possible) and the wire is secured on the skull surface, this results in a

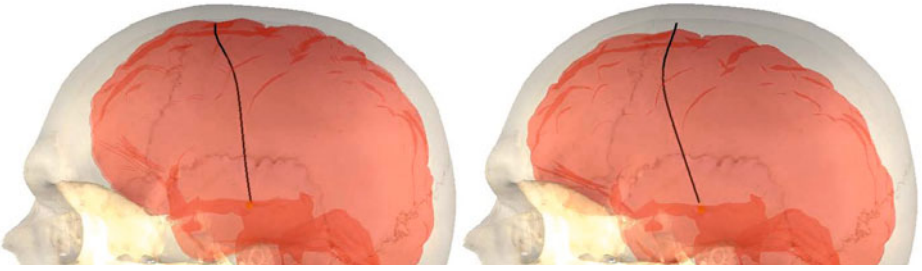


Fig. 4. Screenshot showing the deflection of the right electrode after the cannula removal (left) and after CSF recovery (right).

relative motion of the electrode with respect to its intra-operative location. This leads to a posteroanterior curvature, as illustrate in figure 4 which correlates very well with data from figure 1.

We also quantitatively compared our results (amplitude of electrode migration) with data reported in [13]. In their study, a correlation between the volume of subdural air and upward electrode displacement along its trajectory was determined. The relationship between the displacement D (in mm) and subdural air volume V (in $cm^3 3$) can be empirically described as $D = 2 + 0.08 * V$. We simulated two relatively different amounts of CSF loss, and computed the corresponding values of V (as the difference between the volume of the brain before and after brain shift) and D. We found that for $V = 22cm^3$ a displacement $D = 3mm$ is computed, compared to $3.7mm \pm 2mm$ according to the experimental data [13]. Similarly, for $V = 62cm^3$ a displacement $D = 6.8mm$ is computed, compared to $D = 6.96mm \pm 2.5mm$ according to the experimental law above. This strong correlation between our results and published data illustrates the potential of our method. It is also important to note that the entire simulation of the combined models (from pre- to post-operative stages) only requires a few minutes to be computed.

5 Conclusion

Preliminary results show a good correlation with data reported in the literature, while all our results are obtained using a unique physics-based framework that offers a global approach rather than independently modeling each phenomena. In addition, we account for the interaction between the brain and electrode, and show that our simulations quantitatively correlate with recently reported experimental data. Finally, our computation times remain compatible with future clinical use. Obviously, this is only a first step, and the accuracy of our simulations could certainly be increased by integrating feedback from intraoperative data in the form of additional constraints. This is our next objective, which will also facilitate direct comparison with intra and post-operative results.

Acknowledegment. The authors would like to thank the French Research Agency (ANR) for funding this study through the ACouStiC project.

References

1. Abolhassani, N., Patel, R., Moallem, M.: Needle insertion into soft tissue: A survey. Medical Engineering & Physics 29(4), 413 (2007)
2. Bucki, M., Lobos, C., Payan, Y.: Framework for a low-cost intra-operative image-guided neuronavigator including brain shift compensation. In: IEEE Engineering in Medicine and Biology Society, pp. 872–875 (2007)
3. Clatz, O., Delingette, H., Talos, I.F., Golby, A.J., Kikinis, R., Jolesz, F.A., Ayache, N., Warfield, S.K.: Robust nonrigid registration to capture brain shift from intra-operative mri. IEEE Transactions on Medical Imaging 24(11), 1417–1427 (2005)
4. Deep-Brain Stimulation for Parkinson's Disease Study Group: Deep-brain stimulation of the subthalamic nucleus or the pars interna of the globus pallidus in parkinson's disease. N. Engl. J. Med. 345(13), 956–963 (2001)
5. Duriez, C., Guébert, C., Marchal, M., Cotin, S., Grisoni, L.: Interactive simulation of flexible needle insertions based on constraint models. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5762, pp. 291–299. Springer, Heidelberg (2009)
6. Kim, Y.H., Kim, H.J., Kim, C., Kim, D.G., Jeon, B.S., Paek, S.H.: Comparison of electrode location between immediate postoperative day and 6 months after bilateral subthalamic nucleus stimulation. Acta Neurochir 152(12), 2037–2045 (2010)
7. Kruse, S.A., Rose, G.H., Glaser, K.J., Manduca, A., Felmlee, J.P., Jack Jr., C.R., Ehman, R.L.: Magnetic resonance elastography of the brain. NeuroImage 39(1), 231–237 (2008)
8. Lunn, K.E., Paulsen, K.D., Lynch, D.R., Roberts, D.W., Kennedy, F.E., Hartov, A.: Assimilating intraoperative data with brain shift modeling using the adjoint equations. Medical Image Analysis 9(3), 281–293 (2005)
9. Lurig, C., Hastreiter, P., Nimsky, C., Ertl, T.: Analysis and visualization of the brain shift phenomenon in neurosurgery. In: TCVG Symposium on Visualization (VisSym), pp. 285–290 (1999)
10. Miga, M., Paulsen, K., Hoopes, P., Kennedy Jr., F., Hartov, A., Roberts, D.: In vivo quantification of a homogeneous brain deformation model for updating preoperative images during surgery. Biomedical Engineering 47(2), 266–273 (2000)
11. Miller, K., Wittek, A., Joldes, G.: Biomechanics of the brain for computer-integrated surgery. publishing House of Warsaw University of Technology (2002)
12. Miyagi, Y., Shima, F., Sasaki, T.: Brain shift: an error factor during implantation of deep brain stimulation electrodes. Journal of Neurosurgery 107(5), 989–997 (2007)
13. Munckhof, P.V.D., Contarino, M.F., Bour, L.J., Speelman, J.D., Bie, R.M.A.D., Schuurman, P.R.: Postoperative curving and upward displacement of deep brain stimulation electrodes caused by brain shift. Journal of Neurosurgery 67(1), 49–54 (2010)
14. Wittek, A., Miller, K., Kikinis, R., Warfield, S.K.: Patient-specific model of brain deformation: Application to medical image registration. Journal of Biomechanics 40(4), 919–929 (2007)
15. Zhang, C., Wang, M., Song, Z.: A brain-deformation framework based on a linear elastic model and evaluation using clinical data. Transactions on Biomedical Engineering 58(1), 1–9 (2011)

Hemodynamic Simulation for an Anatomically Realistic Portal System

Harvey Ho[1], Adam Bartlett[2], and Peter Hunter[1]

[1] Bioengineering Institute, University of Auckland, New Zealand
{harvey.ho,p.hunter}@auckland.ac.nz
[2] Department of Surgery, University of Auckland, Auckland, New Zealand
a.bartlett@auckland.ac.nz

Abstract. In this paper we present a computer model that simulates blood flow in the portal system, the vascular network that delivers nutrients and hormones to the liver from other digestion organs. Firstly the vascular geometry of a patient-specific portal system was digitised from a 3D CT image. Then blood flow in this system was solved using a set of 1D partial differential equations coupled with a bifurcation model. Some preliminary results are presented and compared with published ultrasonography and phase contrast MRA data. We further simulate a surgical procedure (portacaval shunting) that connects the portal vein with the inferior vena cava (IVC). The simulation confirms that the high pressure gradient between the portal vein and IVC leads to substantially reduced portal perfusion or even reversed flow in the portal veins, thus makes the shunting graft a flow 'highway' in directing portal flow.

1 Introduction

The portal system is the venous network that starts from the intestine and terminates at capillaries in the liver [2]. This network, together with the splanchnic network which are the veins connecting the spleen and the liver, constitutes a vital pathway for digestion. It delivers nutrients from the intestine and hormones from the pancreas and spleen into the liver, where nutrient metabolism and detoxification further take place. Pathology of the portal system will hamper food digestion and may even cause death. For example, liver cirrhosis leads to an increased resistance to portal flow, and in turn causes portal hypertension. Serious portal hypertension is the major cause of bleeding esophagus varices, which is fatal if not treated urgently [2].

Hemodynamics research of the portal system is the study of blood flow in the system, in both healthy and pathological conditions. The importance of this research can be appreciated from the amount of literature on this subject over the last several decades. Quantitative flow data have been collected from *in vivo* (e.g., in [1,3,4]) and *in vitro* measurements (e.g., in [5]). Computer models have been designed in line with *in vitro* experiments and clinical observations to explain portal circulation phenomena (e.g., in [5]). These models have provided great insights into the physiology of portal and splanchnic circulation, thus aided

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 347–354, 2011.

clinical and physiological research. For example, an analog electronic model was proposed to provide the theoretical basis for a portacaval shunting procedure [5]. These hemodynamic models, however, are mostly zero-dimensional (0D) models that lump the blood vessels into electronic components (e.g., resistors, capacitors) to represent their resistance and compliance.

The advance of medical scanning techniques and increasing computational power have made it possible to study the portal system in great detail. Not only can a large portal network be segmented for structural analysis [6], but the blood flow in a local portal bifurcation can be simulated for functional study [7]. However, blood flow modelling for a large image-based portal/splanchnic network has not yet been reported, to our knowledge.

The purpose of this paper is two fold. First, we present a vascular network construction method that facilitates further flow simulation. Second, we model blood flow in the portal system by solving governing flow equations. Based on this model, we simulate a surgical procedure whereby a shunting graft is employed to connect the portal vein and the inferior vena cava (IVC). We present some preliminary results and compare the results with published *in vivo* data. We also comment on the performance of this model in the Discussion section.

2 Method

2.1 Blood Vessel Digitisation and Medical Imaging

We retrospectively studied a 3D CT image of an anonymous patient. The image spanned the whole torso (Fig. 1(a)) and its voxel resolution was 0.879mm $\times$ 0.879mm $\times$ 0.625mm. The scanning was performed at the portal-venous phase to enhance the portal and hepatic veins (indicated by arrows) via a contrast agent. The objective of this work was to simulate blood flow in these vessels by solving flow equations. To this end we needed to identify the topology of the network, clarify the connectivity between parent/daughter vessels, and determine the radius and length of each blood vessel in the network. We employed an imaging and visualization tool CMGUI (http://www.cmiss.org/cmgui) for this work. The procedure was as follows:

1. Locate the key points (nodes) along the centreline of blood vessels;
2. Measure vessel diameter at each node; and
3. Connect these nodes using one dimensional (1D) finite element elements.

This process is illustrated in Fig. 1(b): the small red dots are the key nodes, which are connected by 1D linear elements, then updated to incorporate vessel radius to mimic blood vessels. The final portal network, plus the IVC, the renal and iliac veins (as part of systemic veins), are shown in Fig. 1(c). The portal veins are colored in gold, and the systemic veins in blue. We also digitised the torso, liver, heart and the spleen as spatial references for these blood vessels. The timeframe for the digitisation process was 15 hours for the portal vasculature, and 16 hours for the organs and the torso. The whole vasculature contains 109 venous segments, with 57 branch (bifurcating or merging) points, 42 inlets and 18 outlets.

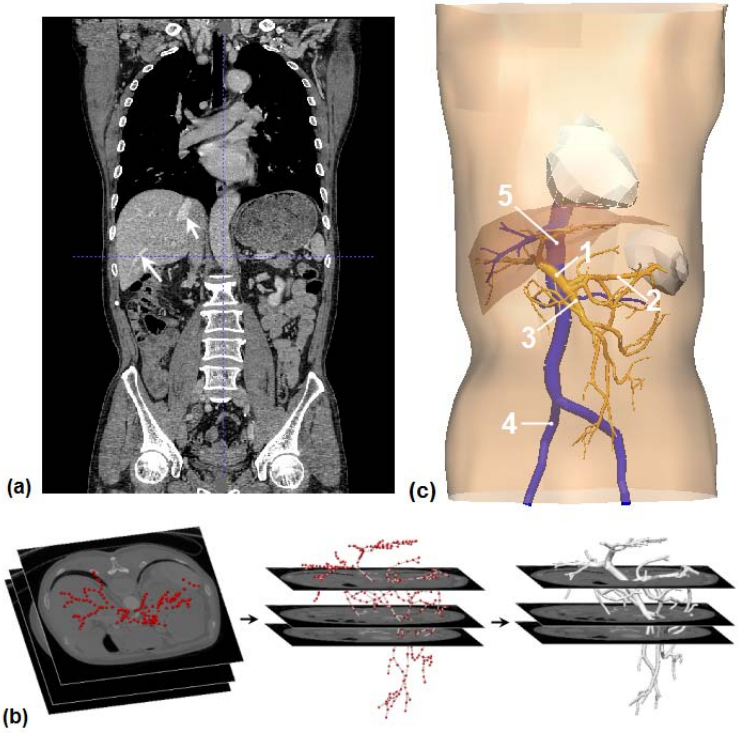


Fig. 1. (a) CT image: the arrows indicate hepatic veins; (b) The process of vessel digitisation; (c) Digitised portal veins and systemic veins. 1 - portal vein; 2 - splenic vein; 3 - superior mesenteric vein (SMV); 4 -iliac vein; 5 - inferior vena cava (IVC).

2.2 Haemodynamics Modelling

We adopted a 1D flow solver to simulate flow in this vasculature. A 1D model essentially works by regarding the longitudinal (or axial) direction as the major flow direction, ignoring flow in other directions [8]. The governing equations are:

$$\frac{\partial R}{\partial t} + V\frac{\partial R}{\partial x} + \frac{R}{2}\frac{\partial V}{\partial x} = 0 \tag{1}$$

$$\frac{\partial V}{\partial t} + (2\alpha - 1)V\frac{\partial V}{\partial x} + 2(\alpha - 1)\frac{V^2}{R}\frac{\partial R}{\partial x} + \frac{1}{\rho}\frac{\partial P}{\partial x} = -2\frac{\upsilon\alpha}{\alpha - 1}\frac{V}{R^2} \tag{2}$$

$$P = \frac{2}{3}\frac{Eh_0}{R_0}\left(\frac{R^2}{R_0^2} - 1\right) \tag{3}$$

where R, P, V are the vessel radius, pressure and flow velocity, respectively. Eqs. (1) and (2) are mass and momentum conservation equations. Eq. (3) is an

empirical vessel wall equation that describes the relationship between transmural pressure and vessel radius. E, h_0, R_0 are Young's modulus, wall thickness and unstressed radius, respectively. α is a constant that defines the axial flow profile. Note, that Eq. (3) only describes the scenario where a vein is expanded (i.e., $P > 0$) and ignores its (partly) collapsed status. The later case is not considered in this work for modelling simplicity's reason.

Eqs. (1)-(3) constitute a closed partial differential equation system for solving hemodynamics in a single vessel. We employed a second-order MacCormack finite difference scheme to solve this system. The spatial and temporal steps were set as 1mm and 0.1 milliseconds respectively for numerical convergence. To simulate flow in a network, we incorporated a bifurcation model which evaluates the flow distribution, velocity and pressure gradient across branches, and therefore the whole tree. We refer the interested reader to the literature [8] for more mathematical details of the bifurcation model.

2.3 Simulation of Portacaval Shunting

Portacaval shunting is a surgical procedure that connects the portal vein with the IVC using an H-graft [1,3,5]. The rationale is that the blood pressure in the IVC is lower than that in the portal vein. The pressure gradient drains portal blood flow into IVC via the shunt to reduce portal pressure and hence relieve the lethal threat of variceal bleeding [2]. The drawback of this procedure, however, is that it can lead to hepatic encephalopathy (i.e., mind-confusion or coma). This is because a large portion of portal blood bypasses the liver, where detoxification takes place, and flows into the brain via blood circulation [2].

The effectiveness of this procedure was investigated by various surgical groups, e.g., by Rypins *et. al.* [1,5]. They found that different shunt sizes led to different clinical outcome. They also used a mathematical model to justify their observations [5]. Based on the pressure and flow data in [5], we can construct a virtual shunt to simulate this procedure, utilizing the vascular network of Fig. 1. In particular, we are able to analyze the portal flow diversion due to the shunting, as shown in the next section.

3 Results

3.1 Flow Simulation for the Portal System

In the first step we simulate flow in the portal system without considering the systemic veins and the shunt. Since the actual measurements for the inlets and outlets of the portal system were not available, we took an empirical approach in the boundary condition (BC) arrangement. This approach is illustrated in Fig. 2. The red dots represent different groups of inlets or outlets that shared the same pressure BC intra-group but there was a pressure gradient between these groups, thus drove flow in the tree. The pressure data in Fig. 2 were adopted from [5].

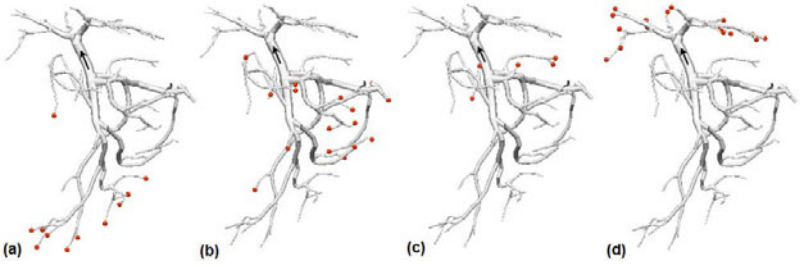


Fig. 2. An empirical approach to pressure boundary condition configuration: (a) inlet group 1: P1=22.5mmHg; (b) inlet group 2: P2=20.6mmHg; (c) inlet group 3: P3=18.75mmHg; (4) outlet group: P4=15mmHg. Note: 1kPa≈7.5mmHg.

Table 1. Comparison of flow velocities in major veins (Unit: mm/s)

Name	Our model	Koslin *et al* [4]	George [7]
Portal vein	71	65	79.9±37.0
SMV	65.4	-	54.4±24.5
Splenic vein	33.6	-	88.5±19.5

As the initial BC, the pressure across the tree was assumed to be the same and there was no flow in the network. Then the blood pressures at the inlets were raised gradually within 0.5 seconds until they reached their respective configured values (as per Fig. 2). The inlet pressures were then held unchanged for another 0.5 seconds to stabilize the simulation. The whole computation took 105 seconds to complete on a desktop computer (Intel Core II, 2.4GHz CPU, 2GB RAM).

Fig. 3(a) and (b) show the pressure and velocity distribution across the portal network. The hierarchical pressure gradient between the inlets and outlets can be appreciated from Fig. 3(a). The pressure profile indicates the time course of inlet pressure BC. The flow velocity distribution, as shown in Fig. 3(b), reflects the fact that the flow velocity is strongly related to vessel diameters. For example, blood flow will be accelerated in a tapering vessel, but decelerated in an expanding vessel.

A comparison was made between the flow velocities calculated from the model and that measured using Duplex ultrasonic measurement [4] and phase contrast MRA [7]. The comparison result, shown in Table 1, indicates that the flow velocities in the portal vein and SMV of our model are consistent with that in literature. The discrepancy of flow velocities in the splenic vein may however be minimized by using a different set of BC and requires further investigation.

3.2 Portacaval Shunt Simulation

Based on the above model, we added systemic veins and the portacaval shunting into the simulation. We analyzed two flow cases: Case 1, an 8mm graft was

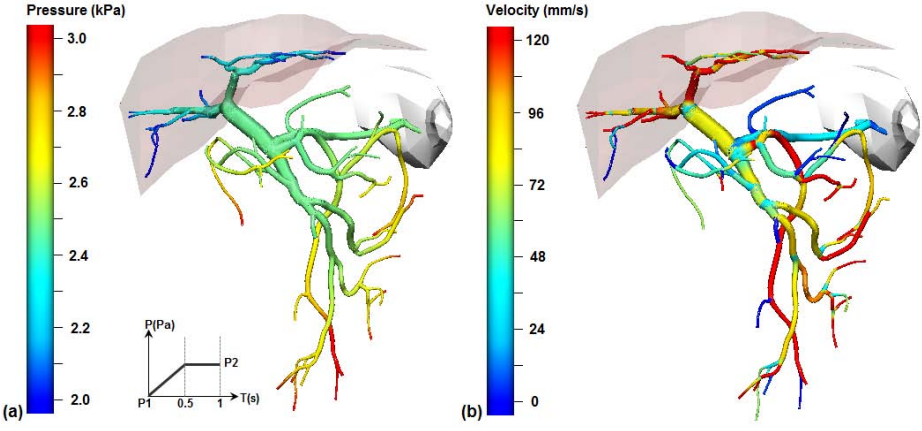


Fig. 3. Blood flow simulation in the portal system: (a) pressure distribution. The pressure profile is the time course of an inlet pressure: it increases from P1 to P2 within 0.5 seconds, then held steady for another 0.5 seconds. The outlet pressure stays at P1 - the pressure gradient between the inlets and outlets drives blood flow; (b) flow velocity distribution: the flow velocity at the portal vein is 71mm/s.

Table 2. Flow distribution in the portal vein and the shunt (Unit: ml/s)

Venous segment	Case 1	Case 2
1	14.46	48.51
2	19.74	31.19
3	5.35	17.73

employed and a prograde flow in the portal vein was sustained; and Case 2, a 10mm shunt was used and the flow in the portal vein was reversed. Both scenarios were observed clinically, e.g., as reported in [1] and [3].

Fig. 4 shows the computational results in the two cases. The flow directions in the shunt, the portal vein pre- and aft-shunt (marked 1, 2, and 3) are indicated by arrows in the sketches. The flow rates ($F = V \cdot \pi R^2$) in these segments are shown in Table 2. In Case 1, $F_1 + F_3 = 19.81(ml/s) \approx F_2$; in Case 2, $F_2 + F_3 = 48.92(ml/s) \approx F_1$. Thus mass conservation in the shunting branch was obeyed.

We observed that a large portion ($14.46 \div 19.74 \approx 73\%$) of portal blood flowed to the shunt in Case 1. While this helps to decompress the portal hypertension, the substantially reduced portal perfusion implies that toxins in the blood were not removed sufficiently. In Case 2 the prograde and retrograde portal merged into shunting flow, thus causing a high flow rate in the graft. Again, this can lead to a high rate of hepatic encephalopathy [1]. The mechanism underlying these phenomena is that the high pressure gradient between the portal vein and the IVC makes the shunting graft a flow 'highway' in directing portal flow.

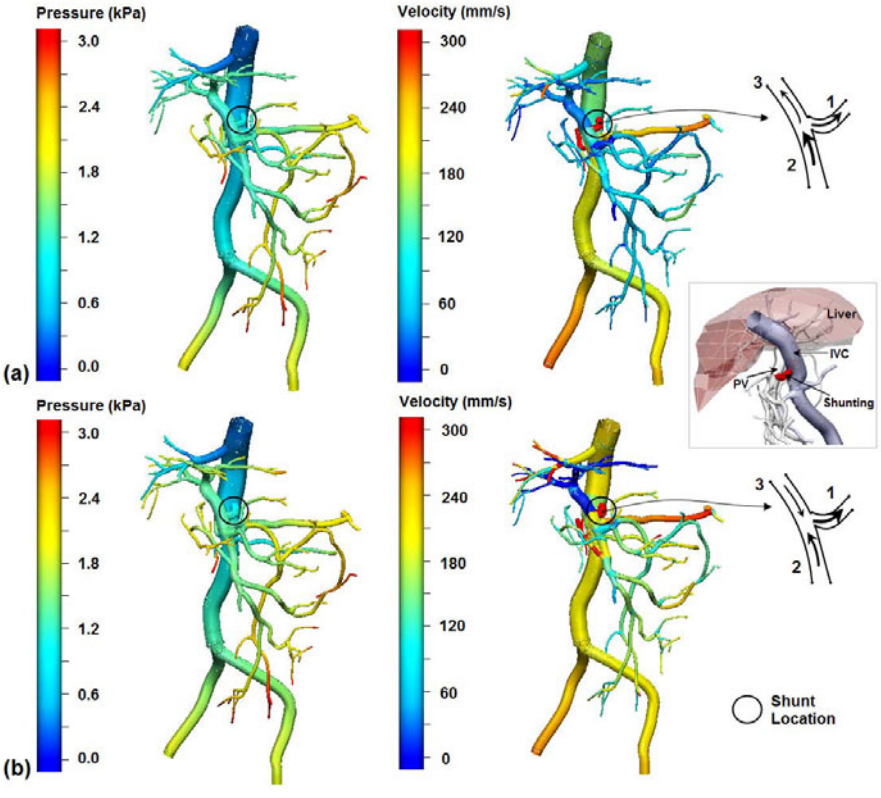


Fig. 4. Flow simulation for the portalcaval shunt: (a) Case 1: prograde flow in the portal vein (using a 8mm graft); (b) Case 2: retrograde flow in the portal vein (using a 10mm graft). The sketches show the flow directions in the portal vein and the shunt.

4 Discussion

This work represents the first portal flow model built on an anatomically accurate portal/splanchnic tree digitised from a 3D scanning image. We employed a 1D solver to calculate the blood flow in this vasculature and used the model to simulate hemodynamics after a surgical procedure that connects the portal vein with the IVC. Some preliminary results were presented and compared with *in vivo* measurements and clinical observations.

The reason we adopted a 1D model instead of a 3D model for the vascular network was due to its computational efficiency. Some other advantages of the 1D model are its ability to capture pressure wave [8] and its flexibility in modelling wall elasticity. However, the convergence of a 1D model is largely affected by boundary conditions. This is particularly true for a large vasculature such as the

one modelled in this paper. While the hierarchical pressure boundary scheme employed in this work can generate a pressure gradient in the network thus the flow, future measured *in vivo* data can further improve the model.

One of the unique flow characteristics of portal circulation is that the hepatic resistance to portal flow is very low. For example, only a 0.5mmHg pressure gradient is required for portal perfusion [5]. Therefore, a small variation of portal pressure due to, e.g., liver cirrhosis, can disturb the portal circulation. These interwoven physiological phenomena require adoption of a combination of flow models, from 0D to 3D, to yield an integrative picture, and thus better interpret and combat portal system diseases.

5 Conclusion

In this paper we simulated blood flow in a portal system that was digitised from a 3D CT image. We used this model to simulate a surgical procedure that connects the portal vein and the IVC via a virtual shunting graft. The simulation confirms the clinical observations that the shunting can induce a substantially reduced portal perfusion or a reversed flow in the portal veins.

References

1. Sarfeh, I.J., Rypins, E.B., Mason, G.R.: A systematic appraisal of portacaval H-graft diameters: Clinical and hemodynamic perspectives. Annals of Surgery 204(4), 356–363 (1986)
2. Boyer, T.D., Wright, T.L., Manns, M.P.: Zakim and Boyer's Hepatology: A textbook of Liver Disease, 5th edn. Saunders (2006)
3. Lafortune, M., Patriquin, H., Pomier, G., Huet, P., Weber, A., Lavoie, P., Blanchard, H., Breton, G.: Hemodynamic changes in portal circulation after portosystemic shunts: use of duplex sonography in 43 patients. Am. J. Roentgenol. 149(4), 701–706 (1987)
4. Koslin, D.B., Berland, L.L.: Duplex Doppler examination of the liver and portal venous system. Journal of Clinical Ultrasound 15(9), 675–686 (1987)
5. Rypins, E.B., Rosenberg, K.M., Sarfeh, I., Houck, J., Conroy, R.M., Milne, N.: Computer analysis of portal hemodynamics after small-diameter portacaval H- grafts: The theoretical basis for partial shunting. Journal of Surgical Research 42(4), 354–361 (1987)
6. Selle, D., Preim, B., Schenk, A., Peitgen, H.: Analysis of vasculature for liver surgical planning. IEEE Transactions on Medical Imaging 21(11), 1344–1357 (2002)
7. George, S.: Hemodynamic investigation of the liver using magnetic resonance imaging and computational fluid dynamics. PhD thesis. Georgia Institute of Technology (2008)
8. Smith, N.P., Pullan, A.J., Hunter, P.J.: An anatomically based model of transient coronary blood flow in the heart. SIAM Journal on Applied Mathematics 62(3), 990–1018 (2001)

Virtual Coiling of Intracranial Aneurysms Based on Dynamic Path Planning

Hernán G. Morales[1], Ignacio Larrabide[1], Minsuok Kim[1], Maria-Cruz Villa-Uriol[1], Juan M. Macho[2], Jordi Blasco[2], Luis San Roman[2], and Alejandro F. Frangi[1]

[1] Center for Computational Imaging & Simulation Technologies in Biomedicine (CISTIB) – Universitat Pompeu Fabra and CIBER-BBN, Barcelona, Spain
[2] Department of Radiology, Hospital Clinic, Barcelona, Spain

Abstract. Coiling is possibly the most widespread endovascular treatment for intracranial aneurysms. It consists in the placement of metal wires inside the aneurysm to promote blood coagulation. This work presents a virtual coiling technique for pre-interventional planning and post-operative assessment of coil embolization procedure of aneurysms. The technique uses a dynamic path planning algorithm to mimic coil insertion inside a 3D aneurysm model, which allows to obtain a plausible distribution of coils within a patient-specific anatomy. The technique was tested on two idealized geometries: an sphere and a hexahedron. Subsequently, the proposed technique was applied in 10 realistic aneurysm geometries to show its reliability in anatomical models. The results of the technique was compared to digital substraction angiography images of two aneurysms.

1 Introduction

Endovascular therapies of intracranial aneurysms are an alternative to the traditional surgical clipping. Among them, coiling is the most popular and common option. This treatment consists in the insertion of biocompatible metal wires inside the aneurysm through a catheter. During treatment, angiographic images are produced to guide the catheter and coils while they navigate inside the vasculature. The amount of inserted coils depends on several factors, including the aneurysm morphology, the selected coils and the operator skills, among others. However, clinicians usually try to insert as many coils as possible to achieve high packing densities (defined as the ratio between the inserted coil and aneurysm volume) [1]. The goal is to induce through the coils a mechanical resistance against the pulsatile blood flow to promote the hemodynamic conditions that triggers blood coagulation [2].

In recent years, computational methods have appeared as an interesting source of information for clinicians to support them during diagnosis, planning or evaluation of aneurysm treatments. A virtual stenting technique [3] and an interactive simulator for virtual coiling [4] are examples of these methods. In this work, we

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 355–362, 2011.

present a virtual coiling technique for virtually treating image-based patient-specific cerebral aneurysm models. The proposed method relies on a dynamic path planning algorithm to insert computational models of the coils.

2 Virtual Coiling Technique

The proposed virtual coiling technique uses a dynamic path planning algorithm to mimic coil insertion. The description of this algorithm is presented in the following section.

2.1 The Dynamic Path Planning Algorithm

A virtual coil model, from here on called *a coil*, is defined as a set of ordered points (see equation 1). Here, the index j denotes the coil currently under insertion, while the index i corresponds to the points of that coil.

$$\mathbb{C}_j \equiv \{\mathbf{p}_j^i \in \mathbb{R}^3 : i = 0...n_j\}, \tag{1}$$

where n_j is the number of segments in the coil j. This number is defined as the floor function of the ratio between the coil length (L_j) and the coil radius (r_j). For each coil to be inserted both, L_j and r_j are provided as input to our algorithm.

To add a new point $\mathbf{p}_j^{i+1}$ to the coil $\mathbb{C}_j$, a set of candidate locations is created as $\mathbb{H}_{j,i} \equiv \{\mathbf{h}_j^i \in \mathbb{R}^3 : ||\mathbf{p}_j^i - \mathbf{h}_j^i|| = r_j\}$. The number of candidate locations ($|\mathbb{H}_{j,i}|$) is a parameter of the algorithm, but in general we have set it to over 150 elements to ensure a proper spatial discretization. The $\mathbf{h} \in \mathbb{H}_{j,i}$ with the lowest potential field $\phi(h)$ is selected and included in $\mathbb{C}_j$ as $\mathbf{p}_j^{i+1}$. This potential field $\phi(h)$ is based on three rules, namely: coiling domain, coil flexibility and coil pull-back, and it is mathematically defined as follows:

$$\phi(h) = \phi_D(h) + \phi_F(h) + \phi_{PB}(h) \ . \tag{2}$$

Coiling Domain: It represents the region where the coil is allowed to move and it is defined as the sum of two potential fields.

$$\phi_D(h) = \phi_A(h) + \phi_C(h) \ . \tag{3}$$

the first term, $\phi_A(h) : \mathbb{R}^3 \rightarrow \{0, 1\}$, is defined by the aneurysm model to be filled with coils. If the candidate location $\mathbf{h}$ is inside the model, $\phi_A(h)$ is set to zero, otherwise to one (see Fig. 1A). The second term, $\phi_C(h) : \mathbb{R}^3 \rightarrow \{[0, \frac{1}{2}), 1\}$, avoids crossings of the coils. Also, it accounts for the distance between coils such that newly inserted coils avoid highly dense coiled areas. The field $\phi_C(h)$ is defined as follows:

$$\phi_C(h) = \begin{cases} 1, & \text{if } \exists \ \mathbf{m} \text{ satisfying } ||\mathbf{h} - \mathbf{m}|| \leq 2 \cdot r_j, \\ \dfrac{r_j}{|\mathbb{M}_{j,i}|} \cdot \displaystyle\sum_m \frac{1}{||\mathbf{h} - \mathbf{m}||}, & \text{otherwise .} \end{cases} \tag{4}$$

where $\mathbf{h} \in \mathbb{H}_{j,i}$ and $\mathbf{m} \in \mathbb{M}_{j,i}$. The set $\mathbb{M}_{j,i}$ (see equation 5) is a subset of all points of the inserted coils within a distance of $\alpha \cdot r_j$ from the last inserted point of $\mathbb{C}_j$, called the coil tip.

$$\mathbb{M}_{j,i} \equiv \{\mathbf{m} \in \bigcup_q \mathbb{C}_q : ||\mathbf{p}_j^i - \mathbf{m}|| \leq \alpha \cdot r_j, q = 1...j\} \ , \tag{5}$$

The Euclidean distance, $\alpha \cdot r_j$, with $\alpha = 4$ was used to limit the size of $\mathbb{M}_{j,i}$, and thereby, to reduce the computational cost of the algorithm. This value of α was set to take into account at least the closest coil points that could potentially produce coil crossings. Figure 1B shows a schematic representation of how $\phi_C(h)$ is evaluated.

Coil Flexibility: It takes into account the deformation of the coil while it is inserted, by comparing the angle θ of the evaluated candidate location with respect to the previous direction $(\mathbf{p}_j^i - \mathbf{p}_j^{i-1})$ as it is visualized in figure 1C. The potential field $\phi_F(h) : \mathbb{R}^3 \rightarrow [0, \frac{1}{2}]$ is defined as:

$$\phi_F(h) = \theta/(2 \cdot \pi) \ , \tag{6}$$

where $\theta = \arccos[(\mathbf{h} - \mathbf{p}_j^i) \cdot (\mathbf{p}_j^i - \mathbf{p}_j^{i-1})/r_j^2]$.

Coil Pull-Back: It is meant to solve the situation when $\phi(h) \geq 1 \ \forall \ \mathbf{h} \in \mathbb{H}_{j,i}$ (becoming a dead-end). In such cases the algorithm pulls back and relocates the coil tip to the position $\mathbf{p}_j^{i-(k+1)}$, being k the number of consecutive dead-ends. Then, a potential field, $\phi_{PB} = 1$ is assigned to the location $\mathbf{p}_j^{i+1}$, and thereby, the dead-end is avoided. An schematic example of the coil pull back is presented in figure 1D.

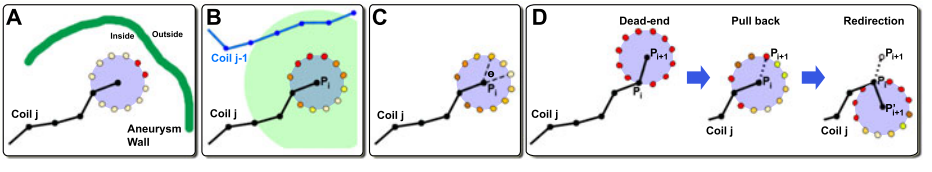


Fig. 1. Two-dimensional schematic description of virtual coiling technique. (A) Rejected candidate locations (red dots) due to $\phi_A = 1$, since their distance to the aneurysm wall is less than one coil diameter. (B) Evaluation of ϕ_C. The set $\mathbb{M}_{j,i}$ contains the coil points inside the green circle. (C) Evaluation of ϕ_F. (D) Dead-end condition, pull back and redirection of the last coil.

Algorithm Initialization: To initialize $\mathbb{C}_j$, a set of initial candidate locations, $\mathbb{S}_j$, is defined around the geometrical center of the aneurysm model within a distance of $\beta \cdot r_j$, with $\beta = 4$. The value of β was set to guarantee a sufficient distance between any two initial coil points. Afterwards, the element $\mathbf{s}$ of $\mathbb{S}_j$ with the lowest $\phi_C(s)+\phi_F(s)$ is selected as the first coil point $\mathbf{p}_j^0$. A user-defined initial direction is set (for instance $\hat{x}$), to calculate $\phi_F(s)$ during the evaluation of the first two coil points $\mathbf{p}_j^0$ and $\mathbf{p}_j^1$.

The proposed approach is summarized in Algorithm 1.

Virtual Coiling Technique based on dynamic path planning;
input : Aneurysm model, number of coils (J), coil radius (r_j) and length (L_j), initial direction and number of candidates.
output: Inserted virtual coils.

begin
for $j = 1$ **to** J **do**
  $n_j = \lfloor (\frac{L_j}{r_j}) \rfloor$;
  Create $\mathbb{S}_j$;
  s' $= \underset{\mathbf{s}\in\mathbb{S}_j}{\text{argmin}}(\phi_C + \phi_F)$
  s' $\rightarrow \mathbb{C}_{j,0}$;
  for $i = 0$ **to** $i = n_j$ **do**
    Create the set $\mathbb{H}_{j,i}$;
    for $\mathbb{H}_{j,i}$ **do**
      h' $= \underset{\mathbf{h}\in\mathbb{H}_{j,i}}{\text{argmin}}(\phi)$
    if $\phi(\boldsymbol{h'}) < 1$ **then**
      h' $\rightarrow \mathbf{p}_j^{i+1}$.
      $i \rightarrow i+1$;
      k=0
    else
      if *there are consecutive dead-ends* **then**
        $k \rightarrow k + 1$;
      $i \rightarrow i - (k + 1)$;
      $\phi_{PB} = 1$ at $\mathbf{p}_j^{i+(k+1)}$;
  Finalize $\mathbb{C}_j$ with $(n_j + 1)$ elements;
  $j \rightarrow j+1$;
end

Algorithm 1. Pseudocode of algorithm for virtual coiling

3 Test on Idealized Geometries

We tested how our method behaves inside idealized geometries. For this two geometries were created (see figure 2): a 2mm-radius sphere, and a 4x4mm^2-base 1.5mm-height hexahedron with a 0.3mm-diameter 0.5mm-height tube above it. Both geometries, sphere and hexahedron, were used to show that the method adapts the shape of the coils to the morphology of the containing geometry, although the hexahedron does not represent a anatomical structure. The small tube on top of the hexahedron was added to force a dead-end and to illustrate how the coil pull-back mechanics operates. This dead-end will appear since the first coil will go straight to the tube end and no chance to bend the coil will be available since the tube diameter would be slightly higher than the coil one.

Five experiments were performed, four of them using the sphere and the last one using the hexahedron. Table 1 summarizes each experiment. The test I and II are meant to see the differences in the coil configuration by changing the coil lengths, maintaining the total coil length (19cm). The objective of tests III and IV is to see the influence of the number of candidate locations and to compare the required computational times against the previous experiments.

Table 1. Experimental set-up on idealized geometries

Test	Idealized Model	Number of coils	Coil diameter [mm]	Coil lengths [cm]	Number of candidate locations	Initial direction
I	Sphere	3	0.25	2-2-15	350	$\hat{z}$
II	Sphere	3	0.25	5-5-9	350	$\hat{z}$
III	Sphere	3	0.25	2-2-15	150	$\hat{z}$
IV	Sphere	3	0.25	5-5-9	150	$\hat{z}$
V	Hexahedron	3	0.25	2-2-20	350	$\hat{z}$

Note that, although clinicians first insert the longest coils, here the shortest ones were placed before to show how the coils are initiated and to have a better idea of the algorithm functioning when the models are almost empty.

3.1 Results on Idealized Models

Figure 2A, B and C present the result of Test I with one, two and the three coils, respectively. The obtained packing density with these coils was 27.3%. The result of Test II is visualized in figure 2D, where a different coil configuration was obtained compared to Test I, although the same high packing density of 27.3% was achieved. Figure 2E and F showed the results of Test II and IV, respectively. As it is visualized, the reduction of candidate locations produces different coil configurations, and the coil roughness increased (sharper angles). The same packing density of 27.3% was obtained with these two experiments.

Test I and II required 230s each to be computed in a Intel Core(TM)2 Quad CPU Q6600 @2.4GHz with 8 GB of RAM. This time was reduced to 93 seconds in Test III and IV, due to the reduction of candidate locations.

The results of Test V are presented in figure 2G to J. To show how the coil pull back works, two iterations are presented in figure 2G, corresponding to the 7^{th} (transparent coil inside the tube) and the 15^th (outside the tube) iteration. When the coil is inside the tube, the algorithm pushes the tip towards the $\hat{z}$ axis. However, once the tip reaches the end of the tube, the algorithm progressively pulls back the tip until it is outside the tube, then the algorithm blocks the tube entrance by adding a potential field $\phi_{PB} = 1$. Therefore, the algorithm selects a candidate location outside the tube, forcing the tip to follow another direction as it is presented with the iteration 15. Figure 2H, I and J present how the algorithm fully inserts the first, second and third coil, respectively. The final packing density inside the hexahedron was 23.7%.

4 Evaluation in Real Geometries

To evaluate the results of our virtual coiling technique on patient-specific aneurysm geometries, two coiled aneurysms were selected. The first case was located at the right middle cerebral artery bifurcation and the second case at the anterior communicating artery. For each of these cases, a three dimensional

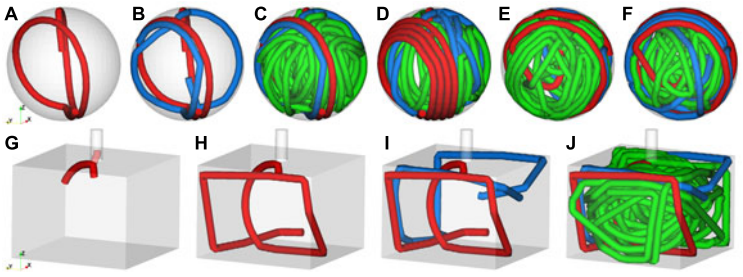


Fig. 2. Virtual coiling on idealized geometries. Test I is presented in (A) after one coil, (B) after two coils and (C) after three coils. (D), (E) and (F) show Test II, III and IV, respectively. Test V is depicted in (G) for two iterations of the first coil, (H) after one coil, (I) after two coils and (J) after three coils.

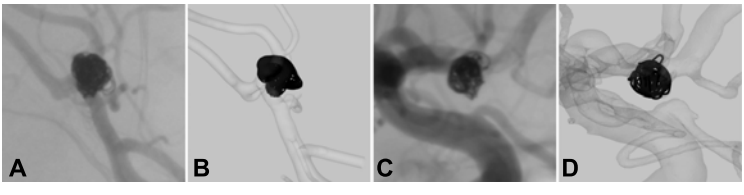


Fig. 3. Comparison of a DSA and a virtual models of case 1, (A) and (B); and case 2, (C) and (D)

rotational angiography (3DRA) image before coiling and a digital subtraction angiography (DSA) image after coiling were acquired. Besides, the treatment information (number of coils and type) was recorded. For case 1, 11 coils were inserted with diameters between 0.245mm and 0.29mm and a total length of 67cm. In case 2, 6 coils were inserted with diameter between 0.245mm and 0.282mm and a total length of 24cm.

Using the two selected cases, the following procedure was applied in each of them. The 3DRA image was segmented using a geodesic active region to get a 3D geometrical representation of the arteries and aneurysm [5]. Subsequently, the aneurysm geometry was extracted and isolated from the rest of the vasculature by placing a surface at its ostium. The virtual coiling technique was applied on the aneurysm model following the treatment procedure chosen for that case.

To compare the outcomes of the proposed technique in these realistic anatomies, a DSA image versus a virtual coiling image of the same case is presented in figure 3. A similar orientation of the 3D models was created, to match the virtual models with the DSA view. This orientation was created by a manual registration of the 3D geometries on the DSA image. In both cases, a similar filled area with coils was observed, although some wires produced particular differences.

Additional Cases: Figure 4 presents 10 additional patient-specific aneurysm models that were virtually treated. To apply our technique, each aneurysm was

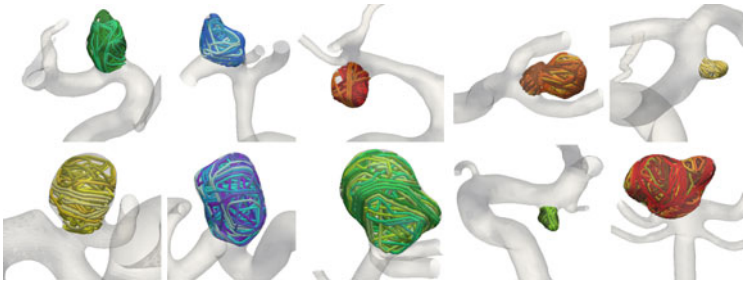


Fig. 4. Additional patient-specific aneurysm models

previously closed by adding a surface at its ostium. As it is visualized, the aneurysm location, morphology and size were different among cases. Besides, neither a DSA image sequence nor treatment data was available for these cases for comparison, and thereby, the presented results are meant for testing and visualization. Moreover, These results show the reliability of the technique when patient-specific models are used. The obtained packing densities were between 19% to 34.5%.

5 Discussion

We have presented a technique to virtually coil intracranial aneurysms, based on a dynamic path planning algorithm to create the devices. This method relies on the facts that the final coil distribution is quasi-random, that currently it is not possible to reproduce exactly the same position and orientation of a real inserted coil inside a patient-specific aneurysm, and that clinicians aim at high packing densities when treating with coils [1].

The goal of this method is not for training, where physics-based models are essential [4]. Instead, we aim to understand the macroscopic behavior of the coil mesh, but considering the geometrical features of each coil. The simplicity of our technique allows reproducing realistic inserted coils and distribution and especially, obtaining similar results compared to physics-based techniques. Moreover, it produces high packing densities in patient-specific (34.5%) and idealized models (27.3%), which has been reported as a key parameter for coiling outcome [1].

The proposed method is fast to compute as it was observed from our results. The number of candidate locations is the most important parameter that impact the computational time. Elevating the number of candidate locations increases the computational time but also enhances the smoothness of the coils. Additionally, the computational time is proportional to the coil length, which clinically depends on the aneurysm size.

The application of our technique is focused on pre- and post-operative planning and diagnosis. For planning, the method calculates the packing density per

coil, allows visualizing the coil distribution inside the aneurysm (angiography-like and 3D) and assessing the coiling outcome from a hemodynamics point of view using computational fluid dynamics solvers. For post-operative diagnosis, it is possible to assess for instance, the influence of different coil diameters or coil configurations on hemodynamics. Besides, the use of fluid-structure interaction methods could help to understand coil compaction phenomenon, which is one of the main causes of aneurysm recanalization after coiling.

6 Conclusion

We have described a technique for the virtual coiling of intracranial aneurysm, which can be used in either idealized geometries or patient-specific anatomies. The technique is based on a dynamic path planning algorithm to insert the coils. A comparison with DSA images for two real coiled aneurysm was performed and similar results were obtained. Additionally, 10 patient-specific aneurysm models were used show the reliability of the method in anatomical models.

Acknowledgments. The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement nr 223920, VPH-NoE project and has been partially funded by the Industrial and Technological Development Center (CDTI) under the CENIT-CDTEAM and CENIT-cvREMOD programs and by the European Commissions project @neurIST (IST-2005-027703). Alejandro F. F. is partially funded by the ICREA-Academia programme.

References

1. Li, M.H., Gao, B.L., Fang, C., Gu, B.X., Cheng, Y.S., Wang, W., Scotti, G.: Angiographic follow-up of cerebral aneurysms treated with Guglielmi detachable coils: an analysis of 162 cases with 173 aneurysms. AJNR Am. J. Neuroradiol. 27(5), 1107–1112 (2006)
2. Wootton, D.M., Ku, D.N.: Fluid Mechanics of Vascular Systems, Diseases, and Thrombosis. Annu. Rev. Biomed. Eng. 1, 299–329 (2003)
3. Larrabide, I., Kim, M., Augsburger, L., Villa-Uriol, M.C., Rufenacht, D., Frangi, A.F.: Fast virtual deployment of self-expandable stents: Method and in vitro evaluation for intracranial aneurysmal stenting. Med. Image Anal. (2010) (in press)
4. Dequidt, J., Marchal, M., Duriez, C., Kerien, E., Cotin, S.: Interactive simulation of embolization coils: Modeling and experimental validation. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 695–702. Springer, Heidelberg (2008)
5. Bogunovic, H., Pozo, J.M., Villa-Uriol, M.C., Majoie, C.B., van den Berg, R., van Andel, H.A., Macho, J.M., Blasco, J., Roman, L.S., Frangi, A.F.: Automated segmentation of cerebral vasculature with aneurysms in 3DRA and TOF-MRA using geodesic active regions: an evaluation study. Med. Phys. 38(1), 210–222 (2011)

Time Domain Simulation of Harmonic Ultrasound Images and Beam Patterns in 3D Using the *k*-space Pseudospectral Method

Bradley E. Treeby[1], Mustafa Tumen[2], and B.T. Cox[2]

[1] College of Engineering and Computer Science, The Australian National University
[2] Department of Medical Physics and Bioengineering, University College London

Abstract. A *k*-space pseudospectral model is developed for the fast full-wave simulation of nonlinear ultrasound propagation through heterogeneous media. The model uses a novel equation of state to account for nonlinearity in addition to power law absorption. The spectral calculation of the spatial gradients enables a significant reduction in the number of required grid nodes compared to finite difference methods. The model is parallelized using a graphical processing unit (GPU) which allows the simulation of individual ultrasound scan lines using a $256 \times 256 \times 128$ voxel grid in less than five minutes. Several numerical examples are given, including the simulation of harmonic ultrasound images and beam patterns using a linear phased array transducer.

Keywords: ultrasound simulation, nonlinear, *k*-space methods, GPU.

1 Introduction

The simulation of ultrasound propagation through biological tissue has many applications. These include the design of ultrasound transducers; the development of new apodization, beamforming, and signal processing techniques; training ultrasonographers to use ultrasound equipment and interpret ultrasound images; medical image registration; and treatment planning [9]. However, simulations based on the conservation equations that govern the propagation of acoustic waves in tissue can be very computationally expensive. This is because the size of the computational domains can equate to hundreds of wavelengths in each spatial dimension. Moreover, established numerical techniques such as the finite element and finite difference methods require on the order of 10 grid nodes per wavelength to achieve acceptable accuracy. This yields a requirement for computational domains with thousands of grid nodes in each spatial dimension. Consequently, 3D simulations can require large amounts of memory and take days or weeks to run, even when distributed computing systems are used [8,4].

To avoid directly solving the fundamental acoustic conservation equations (or the equivalent wave equations), researchers have previously been forced to make a number of simplifying assumptions. For example, for computing the shape of the acoustic field produced by different ultrasound transducers, a parabolic

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 363–370, 2011.

Table 1. Required grid sizes and memory capabilities for ultrasound simulations in three-dimensions using a k-space pseudospectral model. The number of required elements is based on two grid nodes per wavelength using a sound speed of 1500 m/s. The required memory is based on the storage of 20 single-precision matrices (these are used to store the tissue properties, acoustic variables, etc).

Domain Size (cm^3)	Maximum Freq (MHz)	Number of Required Elements	Required Memory (GB)
5^3	1	67^3	0.022
	5	333^3	2.8
	10	667^3	22
10^3	1	133^3	0.18
	5	667^3	22
	10	1333^3	180

approximation is often used [2]. However, this approximation makes it difficult to accurately model the wave-field away from the transducer axis. For the simulation of diagnostic ultrasound images, both ray-tracing and convolution based approaches are frequently used [14,6]. However, these approaches are not normally extended to model multiple reflections or nonlinear effects. The latter is particularly important as modern ultrasound scanners rely on nonlinear wave propagation (in which acoustic waves at lower frequencies produce waves at higher frequency harmonics as they propagate) for tissue harmonic imaging which gives improved image clarity and contrast. There is thus a need for ultrasound simulation tools with less restrictive assumptions but improved computational efficiency.

Here, the k-space pseudospectral method is applied to the simulation of nonlinear ultrasound propagation in biological tissue. The spectral calculation of the spatial derivatives is performed using the fast Fourier transform (FFT) and only requires two grid nodes per acoustic wavelength to achieve acceptable accuracy [10]. This significantly relaxes the requirement for dense computational domains compared to finite difference methods. The use of the FFT also provides a convenient method for parallelization using graphical processing units (GPUs). A summary of the required computing capabilities is given in Table 1; the equivalent finite difference model would require more than two orders of magnitude more elements rendering most of these problems intractable for normal computing systems. Note, for nonlinear simulations, the maximum frequency may be several times higher than the centre frequency of the transducer. In addition to the spectral calculation of spatial derivatives, the accuracy of the finite difference time step is also improved using a k-space adjustment (this makes the temporal discretization scheme exact in the case of linear propagation in homogeneous media [10]). For heterogeneous media, this is similar to using a pseudospectral method with a higher order scheme for the temporal derivative, but has a smaller memory penalty. This approach has previously been used to model linear wave propagation in biomedical photoacoustics [3] and ultrasonics [4].

The numerical solution is based directly on coupled acoustic conservation equations governing nonlinear wave propagation in heterogeneous media. A novel equation of state is used which accounts for material nonlinearity in addition to power law acoustic absorption (biological tissue has been experimentally shown to exhibit absorption characteristics of this form [9]). An accurate model for acoustic absorption is of particular importance for nonlinear wave propagation as the generation of higher frequencies via nonlinearity is delicately balanced with their absorption. Previous full-wave models have included only thermoviscous absorption (which is inaccurate for biological tissue) or a spectrum of relaxation processes [5]. While the latter can account for power law absorption over a limited frequency range, this requires an a priori fitting of an array of relaxation times for each value of absorption and set of simulation parameters.

2 Model Development

The equations governing the nonlinear propagation of compressional acoustic waves in heterogeneous fluid media can be derived from the mass, momentum, and energy conservation laws for continuum mechanics. Under the assumption of a quiescent, isotropic, and inviscid medium in which shear waves can be neglected, the momentum and mass conservation equations can be respectively written as

$$\rho_0 \frac{\partial \mathbf{u}}{\partial t} + \nabla p = -\rho \frac{\partial \mathbf{u}}{\partial t} - \frac{1}{2}\rho_0 \nabla \mathrm{u}^2 \ , \qquad \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho_0 \mathbf{u}) = -\nabla \cdot (\rho \mathbf{u}) \ . \tag{1}$$

Here p and ρ are the acoustic pressure and density, $\mathbf{u}$ is the particle velocity where $\mathrm{u}^2 = \mathbf{u} \cdot \mathbf{u}$, ρ_0 is the ambient (background) density, and only terms up to second order in the acoustic variables have been retained. Following the approach taken by Aanonsen et al. [1], the second order terms which appear on the right hand side in (1) can be re-written in terms of the Lagrangian density via the repeated substitution of the acoustic equations in linearized form. If only cumulative nonlinear effects are important (as is the case for biomedical ultrasound in the absence of microbubbles), the Lagrangian density terms can then be neglected which yields the expressions

$$\frac{\partial \mathbf{u}}{\partial t} = -\frac{1}{\rho_0} \nabla p \ , \qquad \frac{\partial \rho}{\partial t} = -\rho_0 \nabla \cdot \mathbf{u} - \mathbf{u} \cdot \nabla \rho_0 + \frac{1}{\rho_0 c_0^4} \frac{\partial p^2}{\partial t} \ . \tag{2}$$

The final term in the mass conservation equation corresponds to a convective nonlinearity in which the particle velocity contributes to the wave velocity.

Neglecting thermoviscous losses and instead including a phenomenological loss operator to account for arbitrary power law absorption [13], the expansion of the total pressure using a Taylor series about the equilibrium density for a heterogeneous medium yields the equation of state

$$p = c_0^2 \left(\rho + \mathbf{d} \cdot \nabla \rho_0 + \frac{B}{2A} \frac{\rho^2}{\rho_0} - \tau \frac{\partial}{\partial t} \left(-\nabla^2 \right)^{\frac{y}{2}-1} \rho - \eta \left(-\nabla^2 \right)^{\frac{y+1}{2}-1} \rho \right) \ . \tag{3}$$

The five terms within the brackets separately account for linear wave propagation, heterogeneities in the ambient density, material nonlinearity, power law acoustic absorption, and sound speed dispersion. Here c_0 is the isentropic sound speed, $\mathbf{d}$ is the particle displacement vector, and τ and η are the absorption and dispersion proportionality coefficients given by $\tau = -2\alpha_0 c_0^{y-1}$ and $\eta = 2\alpha_0 c_0^y \tan(\pi y/2)$. These operators account for acoustic absorption of the form $\alpha = \alpha_0 \omega^y$ where α_0 is the absorption coefficient prefactor in Np $(\mathrm{rad/s})^{-y}\,\mathrm{m}^{-1}$ and y is the power law exponent [13]. Note, the $\nabla\rho_0$ terms in (2) and (3) cancel to first order and are not included in the discrete equations given below.

The three expressions given in (2) and (3) form a coupled set of equations for the acoustic particle velocity, density, and pressure. It is also possible to combine these expressions into a single second order wave equation for the acoustic pressure. Neglecting higher order absorption, nonlinearity, and heterogeneity terms, this can be written in the form of a modified Westervelt equation

$$\nabla^2 p - \frac{1}{c_0^2}\frac{\partial^2 p}{\partial t^2} - \frac{1}{\rho_0}\nabla\rho_0 \cdot \nabla p + \frac{\beta}{\rho_0 c_0^4}\frac{\partial^2 p^2}{\partial t^2} + \left(\tau \frac{\partial}{\partial t}\left(-\nabla^2\right)^{\frac{y}{2}} + \eta\left(-\nabla^2\right)^{\frac{y+1}{2}}\right) p = 0 \;, \tag{4}$$

where $\beta = 1 + B/2A$ is the coefficient of nonlinearity. An equation of this form without the power law absorption term has previously been used for the simulation of ultrasound images in 2D using the finite difference method [8,7].

Neglecting higher order absorption and nonlinearity effects, the conservation equations in (2) written in discrete form using a k-space pseudospectral method are given by

$$\frac{\partial}{\partial \xi} p^n = \mathbb{F}^{-1}\left\{ ik_\xi \, \kappa \, \mathbb{F}\left\{ p^n \right\}\right\} \;, \tag{5a}$$

$$u_\xi^{n+1} = u_\xi^n - \frac{\Delta t}{\rho_0}\frac{\partial}{\partial \xi} p^n \;, \tag{5b}$$

$$\frac{\partial}{\partial \xi} u_\xi^{n+1} = \mathbb{F}^{-1}\left\{ ik_\xi \, \kappa \, \mathbb{F}\left\{ u_\xi^{n+1} \right\}\right\} \;, \tag{5c}$$

$$\rho_\xi^{n+1} = \rho_\xi^n - \Delta t \rho_0 \frac{\partial}{\partial \xi} u_\xi^{n+1} + \frac{\left(p^n\right)^2 - \left(p^{n-1}\right)^2}{N \, \rho_0 c_0^4} \;. \tag{5d}$$

Here i is the imaginary unit, k_ξ is the wavenumber in the ξ direction, κ is the k-space adjustment where $\kappa = \mathrm{sinc}(c_0 k \Delta t/2)$, $\mathbb{F}$ and $\mathbb{F}^{-1}$ denote the forward and inverse Fourier transform, Δt is the time step, and (5a)-(5d) are repeated for each Cartesian direction in $\mathbb{R}^N$ where $\xi = (x)$ in $\mathbb{R}^1$, $\xi = (x, y)$ in $\mathbb{R}^2$, and $\xi = (x, y, z)$ in $\mathbb{R}^3$. The corresponding equation of state in discrete form is

$$p^{n+1} = c_0^2 \left(\rho^{n+1} + \frac{B}{2A}\frac{1}{\rho_0}\left(\rho^{n+1}\right)^2 - \mathrm{L} \right) \;, \tag{5e}$$

where the total density is given by $\rho = \sum_\xi \rho_\xi$ and the discrete loss term is

$$\mathrm{L} = -\tau \, \mathbb{F}^{-1}\left\{ (k)^{y-2} \, \mathbb{F}\left\{ \rho_0 \sum\nolimits_\xi \frac{\partial}{\partial \xi} u_\xi^{n+1} \right\}\right\} + \eta \, \mathbb{F}^{-1}\left\{ (k)^{y-1} \, \mathbb{F}\left\{ \rho^{n+1} \right\}\right\} \;. \tag{5f}$$

The discrete equations in (5) are iteratively solved using a time step based on the Courant-Friedrichs-Lewy (CFL) number, where $\Delta t = \mathrm{CFL}\,\Delta x / c_{\max}$. A CFL number of 0.3 typically provides a good balance between accuracy and computational speed for weakly heterogeneous media [10]. At each time step, a velocity (or pressure) source can be included by adding the source values to the appropriate voxels within the computational domain. Similarly, the output from the simulation can be obtained by recording the acoustic variables at each time step at particular voxels within the grid. For the simulations presented here, the computational grids were also spatially and temporally staggered to improve accuracy, and an absorbing boundary layer included to prevent waves from wrapping around the domain [12]. The codes were written as an extension to the k-Wave MATLAB toolbox (`http://www.k-wave.org`) and parallelization was achieved by interfacing with the GPU using Accelereyes Jacket. The reported simulation times correspond to execution on an NVIDIA TESLA C2070 which has 448 CUDA cores and 6 GB of onboard memory.

To validate the discrete equations, the one-dimensional propagation of a 5 MPa sinusoidal pressure source through a lossy medium with $\alpha_0 = 0.25$ dB MHz^{-2} cm^{-1} was compared to the analytical expansion provided by Mendousse [5]. The acoustic pressure versus the non-dimensional shock parameter σ for $0 \leq \sigma \leq 3$ is shown in Fig. 1. There is a close agreement between the two solutions illustrating that cumulative nonlinear effects are correctly encapsulated.

3 Harmonic Ultrasound Simulations

To illustrate the applicability of the developed nonlinear k-space model to ultrasound simulation, the beam pattern produced by a linear phased array transducer was investigated. The utilized transducer model was 30 mm wide with 72 rectangular elements 5 mm in length and a kerf width of 0.25 mm. The computational grid including the absorbing boundary layer was $120 \times 60 \times 30$ mm ($512 \times 256 \times 128$ voxels) supporting a maximum frequency of 3.54 MHz. The transducer was driven by a 5 cycle tone burst with a center frequency of 1 MHz and an equivalent single element source pressure of 1 MPa. The beam was electronically focused at 50 mm (with no elevation focusing), all elements were active, and no transmit apodization was used. The simulation was completed in 12.3 minutes (2360 time steps). The resulting beam patterns are shown using a linear plot scale in Fig. 2. These appear visually similar to those produced in 2D by Wojcik using the pseudospectral method [15]. Note, the same simulation performed using a $256 \times 128 \times 64$ voxel grid and 1080 time steps was computed in 51 seconds which demonstrates the excellent scaling characteristics of the model.

As a second example, the same transducer was used to simulate a B-mode ultrasound image of a scattering phantom (see Fig. 3). The ultrasound image was formed from 33 scan lines swept from -32° to 32° using the conventional procedure of transmit and receive beamforming, time gain compensation, envelope detection, log compression, and scan conversion [9]. The total computational grid was $60 \times 60 \times 30$ mm ($256 \times 256 \times 128$ voxels) supporting a maximum

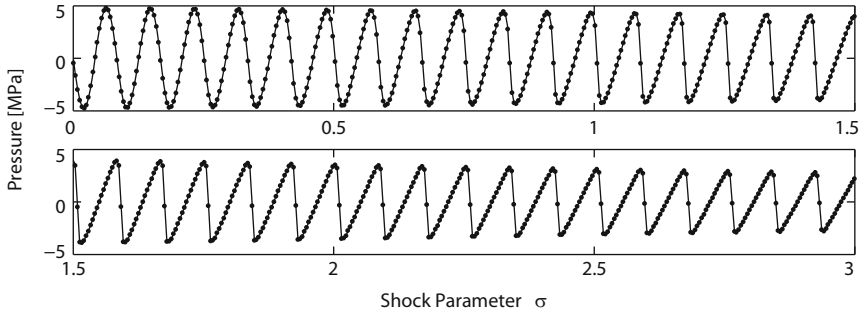


Fig. 1. Evolution of a finite amplitude pressure field in a one-dimensional medium with the shock parameter σ. The k-space solution is shown with a dotted line and the analytical expansion given by Mendousse is shown as a solid line for comparison.

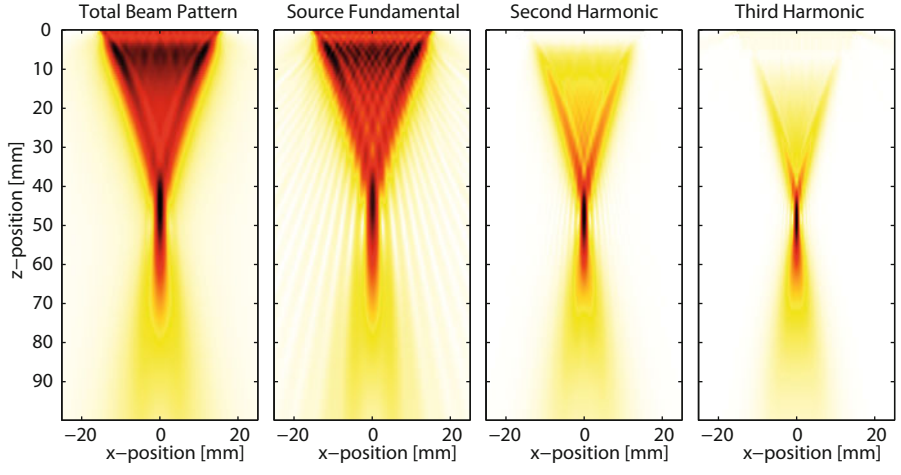


Fig. 2. Normalized beam patterns generated by a linear phased array transducer electronically focused at 50 mm. The total beam pattern is extracted from the integrated frequency spectrum at each position, while the beam pattern at the fundamental and harmonics correspond to the relative spectral amplitudes at these frequencies.

frequency of 3.54 MHz with a time step of 40 ns. The transducer focus was set to 30 mm and the harmonic image was generated using phase inversion. The phantom was created by modulating the mean sound speed and density at each voxel using random Gaussian noise. For contrast, three spherical regions with increased scattering and impedance were defined. Each ultrasound scan line was computed in 4.78 minutes (corresponding to 1700 time steps).

In comparison, Daoud et al. reported simulation times of 2.86 hours per scan line for the same number of total voxels using a linear k-space model on a computer cluster with 2 nodes per scan line and 10 scan lines computed in parallel [4]. While the relative computation time per scan line can be reduced by using

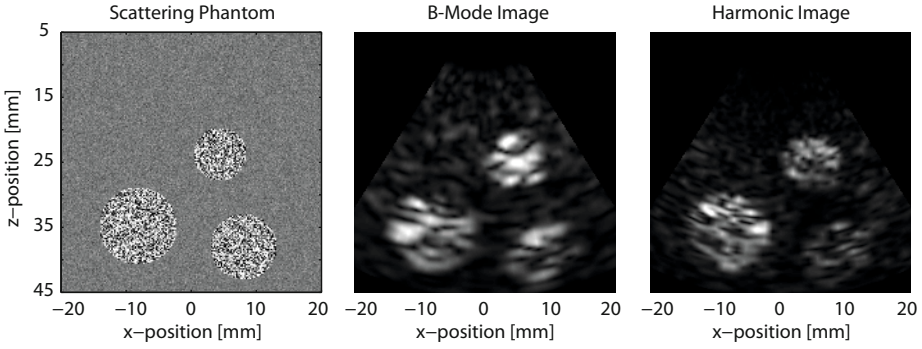


Fig. 3. Simulated B-mode ultrasound images of a scattering phantom. The images look realistic and contain the features and artifacts seen in B-mode images from commercial diagnostic scanners.

more nodes, this also reduces the computational efficiency due to the additional communication required (for 30 nodes the reported efficiency was approximately 90% and for 60 nodes 75%). Thus, to achieve a comparable execution time to that reported here, nearly 100 nodes are required. This illustrates the architectural advantages of using GPUs for spectral methods over conventional computer clusters (the parallel performance of the FFT is discussed in more detail in [11]). It is useful to note that for comparable simulations, finite difference methods yield significantly better performance when normalized by the number of voxels and time steps [8,7]. For example, Karamalis et al. reported simulation times on the order of 20 seconds per scan line for 2D simulations with 2^{22} grid elements and 6000 time steps using a finite difference method accelerated by the GPU [7]. However, for the same accuracy, the finite difference method requires as many as two orders of magnitude more voxels (in 3D) and an order of magnitude more time steps, thus the overall advantage of the k-space method is retained.

4 Summary and Discussion

A fast method for 3D simulations of nonlinear ultrasound propagation in biological tissue has been presented. This incorporates a novel equation of state that includes both nonlinearity and power law acoustic absorption. The spectral calculation of the spatial derivatives significantly reduces the requirement for dense computational domains compared to finite difference and finite element models. The model is applicable to all areas of ultrasound simulation but has particular relevance in treatment planning and exposure limit studies in which an accurate model of acoustic absorption is critical. The model is parallelized using a GPU which yields computational times significantly less than equivalent studies previously reported in the literature. This approach also has the potential to substantially reduce the total time taken to simulate ultrasound images by distributing scan lines across a cluster of GPUs.

Acknowledgments. This work was supported in part by the Australian Research Council/Microsoft Linkage Project LP100100588 and the Engineering and Physical Sciences Research Council, UK. The authors would like to acknowledge Ramtin Shams and Jiří Jaroš for useful discussion, and the NVIDIA Academic Partnership Program for providing the TESLA C2070.

References

1. Aanonsen, S.I., Barkve, T., Tjotta, J.N., Tjotta, S.: Distortion and harmonic generation in the nearfield of a finite amplitude sound beam. J. Acoust. Soc. Am. 75(3), 749–768 (1984)
2. Baker, A.C., Berg, A.M., Sahin, A., Tjotta, J.N.: The nonlinear pressure field of plane, rectangular apertures: Experimental and theoretical results. J. Acoust. Soc. Am. 97(6), 3510–3517 (1995)
3. Cox, B.T., Kara, S., Arridge, S.R., Beard, P.C.: k-space propagation models for acoustically heterogeneous media: Application to biomedical photoacoustics. J. Acoust. Soc. Am. 121(6), 3453–3464 (2007)
4. Daoud, M.I., Lacefield, J.C.: Distributed three-dimensional simulation of B-mode ultrasound imaging using a first-order k-space method. Phys. Med. Biol. 54(17), 5173–5192 (2009)
5. Hamilton, M.F., Blackstock, D.T. (eds.): Nonlinear Acoustics. Acoustical Society of America, Melville (2008)
6. Jensen, J.A., Svendsen, N.B.: Calculation of pressure fields from arbitrarily shaped, apodized, and excited ultrasound transducers. IEEE Trans. Ultrason. Ferroelectr. Freq. Control 39(2), 262–267 (1992)
7. Karamalis, A., Wein, W., Navab, N.: Fast Ultrasound Image Simulation Using the Westervelt Equation. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6361, pp. 243–250. Springer, Heidelberg (2010)
8. Pinton, G.F., Dahl, J., Rosenzweig, S., Trahey, G.E.: A heterogeneous nonlinear attenuating full-wave model of ultrasound. IEEE Trans. Ultrason. Ferroelectr. Freq. Control 56(3), 474–488 (2009)
9. Szabo, T.L.: Diagnostic Ultrasound Imaging. Elsevier, Burlington (2004)
10. Tabei, M., Mast, T.D., Waag, R.C.: A k-space method for coupled first-order acoustic propagation equations. J. Acoust. Soc. Am. 111(1), 53–63 (2002)
11. Tillett, J.C., Daoud, M.I., Lacefield, J.C., Waag, R.C.: A k-space method for acoustic propagation using coupled first-order equations in three dimensions. J. Acoust. Soc. Am. 126(3), 1231–1244 (2009)
12. Treeby, B.E., Cox, B.T.: k-Wave: MATLAB toolbox for the simulation and reconstruction of photoacoustic wave fields. J. Biomed. Opt. 15(2), 021314 (2010)
13. Treeby, B.E., Cox, B.T.: Modeling power law absorption and dispersion for acoustic propagation using the fractional Laplacian. J. Acoust. Soc. Am. 127(5), 2741–2748 (2010)
14. Wein, W., Brunke, S., Khamene, A., Callstrom, M.R., Navab, N.: Automatic CT-ultrasound registration for diagnostic imaging and image-guided intervention. Med. Image Anal. 12(5), 577–585 (2008)
15. Wojcik, G.L., Mould, J., Ayter, S., Carcione, L.M.: A study of second harmonic generation by focused medical transducer pulses. In: IEEE Ultrasonics Symposium Proceedings (Cat. No. 98CH36102), pp. 1583–1588 (1998)

Ultrasound Elastography Using Three Images

Hassan Rivaz, Emad M. Boctor, Michael A. Choti, and Gregory D. Hager

Johns Hopkins University

Abstract. Displacement1 estimation is an essential step for ultrasound elastography and numerous techniques have been proposed to improve its quality using *two* frames of ultrasound RF data. This paper introduces a technique for calculating a displacement field from *three* frames of ultrasound RF data. To this end, we first introduce constraints on variations of the displacement field with time using mechanics of materials. These constraints are then used to generate a regularized cost function that incorporates amplitude similarity of three ultrasound images and displacement continuity. We optimize the cost function in an expectation maximization (EM) framework. Iteratively reweighted least squares (IRLS) is used to minimize the effect of outliers. We show that, compared to using two images, the new algorithm reduces the noise of the displacement estimation. The displacement field is used to generate strain images for quasi-static elastography. Phantom experiments and *in-vivo* patient trials of imaging liver tumors and monitoring thermal ablation therapy of liver cancer are presented for validation.

1 Introduction

Displacement, motion or time delay estimation in ultrasound images is an essential step in numerous medical imaging tasks including the rapidly growing field of imaging the mechanical properties of tissue [1]. In this work, we perform displacement estimation for quasi-static ultrasound elastography [1], which involves deforming the tissue slowly with an external mechanical force, imaging the tissue during the deformation, and performing displacement estimation using the images. More specifically, we focus on real-time freehand palpation elastography [2–7] where the external force is applied by simply pressing the ultrasound probe against the tissue. Ease of use, real-time performance and providing invaluable elasticity images for diagnosis and guidance/monitoring of surgical operations are the key factors that have led to its successful commercialization.

A typical ultrasound frame rate is 20-60 fps. As a result, an entire series of ultrasound images are freely available during the tissue deformation. Multiple ultrasound images have been used before to obtain strain images of highly compressed tissue by accumulating the intermediate strain images, and to obtain persistently high quality strain images by performing weighted averaging of the strain images [8–10]. Accumulating and averaging strain images increases their signal to noise ratio (SNR) and contrast to noise ratio (CNR). However, these techniques are susceptible to drift, a problem with any sequential tracking system

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 371–378, 2011.

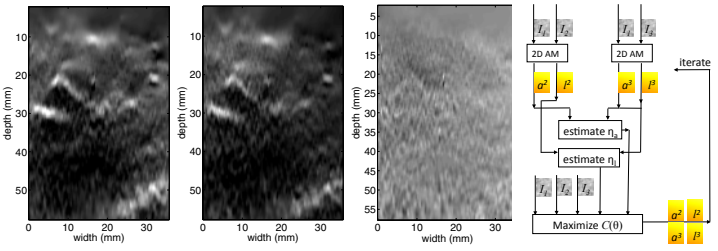


Fig. 1. Left: *in-vivo* images of liver. First and second (S_1 and S_2 from left) are two strain fields calculated from I_1 and I_2, and from I_2 and I_3 respectively. S_1 & S_2 look "similar". Third image is $S_1 - \eta S_2$ for $\eta = 1.1$. The strain range in the first two images is 0 to 0.6%, and in the third image is ±0.3%. Right shows the ElastMI algorithm.

[11]. In addition, these techniques do not exploit additional images to *improve displacement estimation*, which has many applications besides strain estimation. Time series of ultrasound data has also been used to characterize tissue [12] and improve elasticity reconstruction [13] and viscoelastic parameters [14, 15].

Figure 1 shows *two consecutive strain images* calculated from *three ultrasound images* using the 2D analytic minimization (AM) method [16]. Our motivation is to utilize the similarity of these two images to calculate a low variance displacement field from three images. The contributions of this work are: (1) introducing constraints on variation of the motion fields based on similarities of strain images through *time*; (2) proposing an EM algorithm to solve for motion fields using three images, and (3) reporting clinical tstudies of ablation guidance/monitoring, with data collection corresponding to before, during and after ablation.

The rest of this paper is summarized as follows. We first introduce the Elastography using Multiple Images (ElastMI) algorithm for tissue displacement estimation, which minimizes a cost function that incorporates data obtained from three images and exploits mechanical constraints. The estimated low variance displacement field can be used in numerous applications in imaging mechanical properties of tissue; we use it for generating strain images by calculating its spatial derivative. We use phantom and *in-vivo* clinical studies to compare ElastMI versus the recently developed elastography technique of 2D AM (code available online at `www.cs.jhu.edu/~rivaz`) [16].

2 ElastMI: Elastography Using Multiple Images

We have a set of $p = 3$ images I_k, $k = 1 \cdots 3$, each of size $m \times n$. Let the 2D displacement field $\mathbf{d}^k = (a^k, l^k)$ denote the displacement between I_k and I_1, where a refers to the axial (i.e. in the direction of the ultrasound beam) and l to the lateral (i.e. perpendicular to the beam and in the imaging plane) directions. By the choice of reference $\mathbf{d}^1 = 0$. Note that we set I_1 as the reference image to simplify the notation. However, in our implementation we always take the middle image (i.e. I_2) as the reference. Our goal is to calculate a high quality $\mathbf{d}^2$ by utilizing all three images in a group-wise approach.

It is well known that many tissue types display linear strain-stress relation in the 0 to 5% range (see [1] for example). In a freehand palpation elastography setup with ultrasound acquisition rate of 20 fps or more, taking three consecutive images as I_1, I_2, I_3 corresponds to strain values of less than 1% and therefore the linearity assumption is valid. Using this property and some simplifying assumptions, it can be shown that the ratio of the strain and displacement fields in different times is a constant value, i.e. *strain images are similar up to a scale* as in Figure 1. We denote the scale factor by $\boldsymbol{\eta} = (\eta_a, \eta_l)$, and allow it to slightly change spatially to account for small nonlinearities in the tissue. As such, η_a and η_l are themselves scale fields in the axial and lateral directions each of size $m \times n$. Using this notation we have $a^3 = \eta_a . * a^2$ and $l^3 = \eta_l . * l^2$ where $.*$ denotes point-wise multiplication.

Let $\boldsymbol{\theta}$ contain all the displacement unknowns $\mathbf{d}^2$ and $\mathbf{d}^3$. The MAP estimate of $\boldsymbol{\theta}$ is obtained by maximizing its posterior probability

$$\Pr(\boldsymbol{\theta} \mid I_1, I_2, I_3) \propto \Pr(I_1, I_2, I_3 \mid \boldsymbol{\theta}) \Pr(\boldsymbol{\theta}) \tag{1}$$

where we have ignored the normalization denominator. The data term is calculated as $\Pr(I_1, I_2, I_3 \mid \boldsymbol{\theta}) = \Sigma_{\boldsymbol{\eta}} \Pr(I_1, I_2, I_3, \boldsymbol{\eta} \mid \boldsymbol{\theta})$. The summation over the latent variable $\boldsymbol{\eta}$ makes the optimization problem intractable. We therefore use Expectation Maximization (EM) to make the problem tractable as following.

1. Initialize: find an estimate for $\boldsymbol{\theta}$ by applying the 2D AM method [16] to two pairs of images (I_1,I_2) and (I_1,I_3) independently.
2. E-step: find an estimate for $\boldsymbol{\eta}$ using $\boldsymbol{\theta}$ (details below).
3. M-step: update $\boldsymbol{\theta}$ with the current estimate of $\boldsymbol{\eta}$ (details below).
4. Iterate between 2 and 3 until convergence.

The algorithm is shown in Figure 1 right. Note that unlike the traditional EM which maximizes $\Pr(I_1, I_2, I_3 \mid \boldsymbol{\theta})$, we maximize the posterior probability of $\boldsymbol{\theta}$ (Equation 1). Steps 2 and 3 are elaborated below.

Calculating $\boldsymbol{\eta}$ from $\boldsymbol{\theta}$ Using Least Squares. At each sample (i, j) in the displacement field $\mathbf{d}^2_{i,j}$, $i = 1 \cdots m$, $j = 1 \cdots n$ take a window of size $m_w \times n_w$ centered at the sample (m_w and n_w are in the axial and lateral directions respectively and both are odd numbers). Stack the axial and lateral components of $\mathbf{d}^2_{i,j}$ that are in the window in two vectors $\mathbf{a}^2_{i,j}$ and $\mathbf{l}^2_{i,j}$, each of length $m_w \times n_w$. Similarly, generate $\mathbf{a}^3_{i,j}$ and $\mathbf{l}^3_{i,j}$ using $\mathbf{d}^3$. Note that since both displacement fields $\mathbf{d}^2_{i,j}$ and $\mathbf{d}^3_{i,j}$ are calculated with respect to samples on I_1, they correspond to the same sample (i, j). We first calculate the axial component $\eta_{(i,j),a}$ ($\boldsymbol{\eta}_{(i,j)} = (\eta_{(i,j),a}, \eta_{(i,j),l})$). Discarding the spatial information in $\mathbf{a}^2_{i,j}$ and $\mathbf{a}^3_{i,j}$, we can average the two vectors into two scalers $\bar{\mathbf{a}}^2_{i,j}$ and $\bar{\mathbf{a}}^3_{i,j}$ and simply calculate $\eta_{(i,j),a} = \bar{\mathbf{a}}^3_{i,j} / \bar{\mathbf{a}}^2_{i,j}$. However, a more elegant way which also takes into account the spatial information is by calculating the least squares solution to the following over-determined problem (superscript T denotes transpose).

$$\mathbf{a}^2_{i,j} \eta_{(i,j),a} = \mathbf{a}^3_{i,j} \quad \text{giving} \quad \eta_{(i,j),a} = \frac{\mathbf{a}^{2T}_{i,j} \mathbf{a}^3_{i,j}}{\mathbf{a}^{2T}_{i,j} \mathbf{a}^2_{i,j}}, \tag{2}$$

which is what we use in our implementation. To calculate the ratio of the lateral displacement fields $\eta_{(i,j),l}$, we take into account possible lateral slip of the probe, which results in a rigid-body-motion. The rigid-body-motion can be simply calculated by averaging the lateral displacement in $\mathbf{d}^2_{i,j}$ and $\mathbf{d}^3_{i,j}$ in the entire image, and calculating the difference between these two average lateral displacements. The lateral scaling factor $\eta_{(i,j),l}$ can be calculated using an equation similar to 2 where the axial displacement $\mathbf{a}_{i,j}$ is replaced with the lateral displacements $\mathbf{l}_{i,j}$. However, we use the following approach which results in a better estimate for $\eta_{(i,j),l}$. The lateral strain ϵ_l is simply $\nu\epsilon_a$ where ν is an unknown Poisson's ratio. Since ν has a small dynamic range in soft tissue and since the difference between the two displacement maps $\mathbf{d}^2$ and $\mathbf{d}^3$ is small, we can assume that ν does not vary from $\mathbf{d}^2$ to $\mathbf{d}^3$. Therefore, $\eta_{(i,j),l} = \eta_{(i,j),a}$. This gives better estimate for $\eta_{(i,j),l}$ since axial displacement estimation is more accurate [16].

Calculating $\boldsymbol{\theta}$ by Maximizing Its Posterior Probability. To analytically solve the MAP estimate of $\boldsymbol{\theta}$, we assume that the data is independent and that the noise model is Gaussian. Although not completely held in real images, these assumptions are also the foundation behind sum of square difference and correlation based elastography methods, which have been extensively shown to produce reliable results. With these assumptions, the robust MAP estimate for $\boldsymbol{\theta}$ can be obtained by minimizing the following cost function

$$\begin{aligned} C(\boldsymbol{\theta}) = &\sum_{i=1}^{m} w_{12,i}\left(I_1(\mathbf{x}_i) - I_2(\mathbf{x}_i + \hat{\mathbf{d}}_i^2) - \delta\mathbf{d}_i^{2T}\nabla I_2(\mathbf{x}_i + \hat{\mathbf{d}}_i^2)\right)^2 + \\ &\sum_{i=1}^{m} w_{13,i}\left(I_1(\mathbf{x}_i) - I_3(\mathbf{x}_i + \eta_{i,a}\hat{\mathbf{d}}_i^2) - \eta_{i,a}\delta\mathbf{d}_i^{2T}\nabla I_3(\mathbf{x}_i + \eta_{i,a}\hat{\mathbf{d}}_i^2)\right)^2 + \\ &\sum_{i=1}^{m}(\mathbf{d}_i^2 - \mathbf{d}_{i-1}^2)^T A(\mathbf{d}_i^2 - \mathbf{d}_{i-1}^2) \end{aligned} \tag{3}$$

where $\hat{\mathbf{d}}_i^2$ is the estimate obtained using 2D AM, $\delta\mathbf{d}_i^2 = \mathbf{d}_i^2 - \hat{\mathbf{d}}_i^2$ is the update in the displacement that we are looking for, $A = diag(\alpha, \beta)$ is a 2×2 diagonal matrix with tunable regularization weights (α, β) that we adjust manually in this work, and ∇ denotes the gradient operator. Robustness is achieved using IRLS through weights $w_{12,i}$ and $w_{13,i}$ which are calculated as following

$$w_{1k,i} = w(I_1(\mathbf{x}_i) - I_k(\mathbf{x}_i + \hat{\mathbf{d}}_i^k)), \text{for } k = 2, 3, \text{and } w(r_i) = \begin{cases} 1 & |r_i| < T \\ \frac{T}{|r_i|} & |r_i| > T \end{cases} \tag{4}$$

where T is a tunable parameter which determines the residual level for which sample i can be treated as outlier. A small T will treat many samples as outliers.

Setting the derivative of C w.r.t. the axial ($\delta a_i^2 = \delta\mathbf{d}_{i,a}^2$) and lateral ($\delta l_i^2 = \delta\mathbf{d}_{i,l}^2$) components of $\delta\mathbf{d}_i^2$ for $i = 1 \cdots m$ to zero and stacking the $2m$ unknowns in $\delta\mathbf{d}^2 = \left[\delta a_1^2\ \delta l_1^2\ \delta a_2^2\ \delta l_2^2 \cdots \delta a_m^2\ \delta l_m^2\right]^T$ and the $2m$ initial estimates in $\hat{\mathbf{d}}^2 = \left[\hat{a}_1^2\ \hat{l}_1^2\ \hat{a}_2^2\ \hat{l}_2^2 \cdots \hat{a}_m^2\ \hat{l}_m^2\right]^T$ we obtain the linear system of size $2m$:

Table 1. The SNR and CNR of the strain images of Figure 2

	Axial, 2D AM	Axial, ElastMI	Lateral, 2D AM	Lateral, ElastMI
SNR	11.11	12.64	6.06	6.63
CNR	8.48	9.63	2.96	3.39

$$(\mathscr{I}' + \mathcal{D})\delta\mathbf{d}^2 = \boldsymbol{r} - \mathcal{D}\hat{\mathbf{d}}^2, \quad \mathcal{D} = \begin{bmatrix} \alpha & 0 & -\alpha & 0 & 0 & 0 & \cdots & 0 \\ 0 & \beta & 0 & -\beta & 0 & 0 & \cdots & 0 \\ -\alpha & 0 & 2\alpha & 0 & -\alpha & 0 & \cdots & 0 \\ 0 & -\beta & 0 & 2\beta & 0 & -\beta & \cdots & 0 \\ 0 & 0 & -\alpha & 0 & 2\alpha & 0 & \cdots & 0 \\ \vdots & & & & & & \ddots & \\ 0 & 0 & 0 & \cdots & -\alpha & 0 & \alpha & 0 \\ 0 & 0 & 0 & \cdots & 0 & -\beta & 0 & \beta \end{bmatrix}, \tag{5}$$

where $\mathscr{I}'$ is a symmetric tridiagonal matrix with 2×2 matrices $\mathfrak{I}'$ in its diagonal:

$$\mathscr{I}' = diag(\mathfrak{I}'^2(1) \cdots \mathfrak{I}'^2(m)),$$

$$\mathfrak{I}'^2(i) = \begin{bmatrix} (w_{12,i} + w_{13,i}{\eta_{i,a}}^2){I'_{1,a}}^2 & (w_{12,i} + w_{13,i}\eta_{i,a}\eta_{i,l})I'_{1,a}I'_{1,l} \\ (w_{12,i} + w_{13,i}\eta_{i,a}\eta_{i,l})I'_{1,a}I'_{1,l} & (w_{12,i} + w_{13,i}{\eta_{i,l}}^2){I'_{1,l}}^2 \end{bmatrix} \tag{6}$$

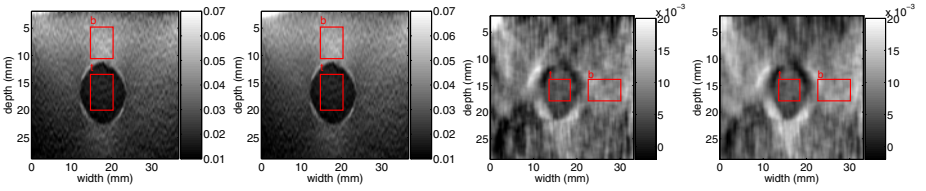
where I'_2 and I'_3 are calculated respectively at $(\mathbf{x}_i + \hat{\mathbf{d}}_i^2)$ and at $(\mathbf{x}_i + \boldsymbol{\eta}_i. * \hat{\mathbf{d}}_i^2)$, superscript $'$ indicates derivative and subscript a and l determine whether the derivation is in the axial or lateral direction, and $\boldsymbol{r}$ is a vector of length $2m$ with elements:

$$\begin{aligned} i \text{ even}: \; r_i &= w_{12,i}I'_{1,a}(\mathbf{x}_i)\left[I_1(\mathbf{x}_i) - I_2(\mathbf{x}_i + \hat{\mathbf{d}}_i^2)\right] + \\ &\quad w_{13,i}\boldsymbol{\eta}_i. * I'_{1,a}(\mathbf{x}_i)\left[I_1(\mathbf{x}_i) - I_3(\mathbf{x}_i + \boldsymbol{\eta}_i. * \hat{\mathbf{d}}_i^2)\right] \\ i \text{ odd}: r_i &= w_{12,i}I'_{1,l}(\mathbf{x}_i)\left[I_1(\mathbf{x}_i) - I_2(\mathbf{x}_i + \hat{\mathbf{d}}_i^2)\right] + \\ &\quad w_{13,i}\boldsymbol{\eta}_i. * I'_{1,l}(\mathbf{x}_i)\left[I_1(\mathbf{x}_i) - I_3(\mathbf{x}_i + \boldsymbol{\eta}_i. * \hat{\mathbf{d}}_i^2)\right]. \end{aligned} \tag{7}$$

The coefficient matrix in Equation 5 is pentadiagonal and symmetric. As such, it can be solved in $8m$ operations, significantly less than $(2m)^3/3$ required for solving a full system. For all the results presented in this work, the EM algorithm is iterated once.

3 Results of Phantom Experiments and Patient Trials

RF data is acquired from an Antares Siemens system (Issaquah, WA) at the center frequency of 6.67 MHz with a VF10-5 linear array at a sampling rate



(a) 2D AM axial (b) ElastMI axial (c) 2D AM lateral (d) ElastMI lateral

Fig. 2. Strain images of the CIRS phantom with the target and background windows (for calculation of SNR and CNR). No Kalman filter [16] is used to ease the comparison.

Table 2. The CNR of the strain images of first, second and third patient trials (images of second patient are shown in Figure 3). P1, P2 and P3 respectively correspond to patients 1, 2 and 3. $2(\bar{s}_b - \bar{s}_t)^2$ and $\sigma_b^2 + \sigma_t^2$ indicate contrast and noise respectively.

		before ablation		during ablation		after ablation	
		2D AM	ElastMI	2D AM	ElastMI	2D AM	ElastMI
P1	$10^4 \times 2(\bar{s}_b - \bar{s}_t)^2$	-	-	-	-	2.18	2.22
	$10^4 \times (\sigma_b^2 + \sigma_t^2)$	-	-	-	-	0.108	0.083
	CNR $= \sqrt{\frac{2(\bar{s}_b - \bar{s}_t)^2}{\sigma_b^2 + \sigma_t^2}}$	-	-	-	-	4.49	5.17
P2	$10^4 \times 2(\bar{s}_b - \bar{s}_t)^2$	0.45	0.89	-	-	2.08	2.15
	$10^4 \times (\sigma_b^2 + \sigma_t^2)$	0.0036	0.0045	-	-	0.204	0.142
	CNR $= \sqrt{\frac{2(\bar{s}_b - \bar{s}_t)^2}{\sigma_b^2 + \sigma_t^2}}$	11.16	14.05	-	-	3.19	3.89
P3	$10^4 \times 2(\bar{s}_b - \bar{s}_t)^2$	0.235	0.234	0.0745	0.1716	4.85	4.82
	$10^4 \times (\sigma_b^2 + \sigma_t^2)$	0.0045	0.0036	0.0091	0.0161	0.204	0.171
	CNR $= \sqrt{\frac{2(\bar{s}_b - \bar{s}_t)^2}{\sigma_b^2 + \sigma_t^2}}$	7.22	8.01	2.87	3.26	4.87	5.31

of 40 MHz. An elastography phantom (CIRS elastography phantom, Norfolk, VA) is compressed axially in two steps using a linear stage, and three images are acquired. Resulting strain images are shown in Figure 2. The unitless metric signal to noise ratio (SNR $= \frac{\bar{s}_b}{\sigma_b}$) and contrast to noise ratio (CNR $= \sqrt{\frac{2(\bar{s}_b - \bar{s}_t)^2}{\sigma_b^2 + \sigma_t^2}}$) [1] of the ElastMI algorithm are shown in Table 1 (The SNR is only calculated for the background window). Comparing to the 2D AM, the ElastMI algorithm improves the SNR by approximately 14% and the CNR by approximately 11%. The high quality of the lateral strain image, compared to state of the art strain imaging techniques, is visually noticeable.

In the clinical studies, RF data was acquired from ablation therapy of three patients with liver cancer using the Siemens Antares ultrasound machine in the following way: for the first patient only after ablation, for the second patient before and after ablation, and for the third patient before, during and after ablation. The ablation was administered using the RITA Model 1500 XRF generator (Rita Medical Systems, Fremont, CA). Tissue was simply compressed freehand at a frequency of approximately 1 compression per 2 sec with the ultrasound probe without any attachment, and the strain images are generated offline.

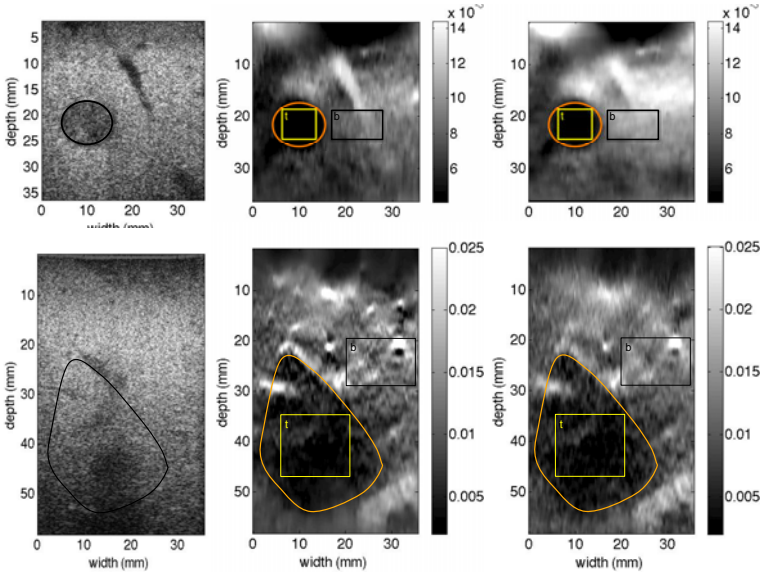


Fig. 3. Axial strain images of the second *in-vivo* patient trial corresponding to before (top row) and after (bottom row) ablation. The first, second and third columns are respectively B-mode, 2D AM strain and ElastMI strain images. The cancer tumor in the top row, and the ablated lesion in the bottom row are delineated. The CNR between the target and background (marked by t & b) windows are given in Table 2.

Results of the second patient trial are shown in Figure 3. Considering the numerous sources of noise in the clinical data, the high contrast of the tumor (top row) and the ablated lesion (bottom row) in the strain images make ElastMI a promising tool for both finding the tumor and monitoring the ablation. It should be noted that elastographic analysis of the ablated lesion is known to be challenging due to high temperatures which significantly degrade the quality of ultrasound data (mainly because of the air bubbles). Table 2 summarizes the CNR, as well as noise and contrast values, in the patient trials obtained using 2D AM and ElastMI methods. In the six cases presented in this table (two before ablation, one during ablation and three after ablation), the average increase in the CNR achieved using ElastMI compared to 2D AM is 17%.

4 Conclusions

In this work, we proposed to utilize three ultrasound images to calculate high quality displacement fields. We neglected the dynamics of tissue motion and assumed a static model for tissue mechanics, which is valid in the quasi-static elastography. Using this model and assuming tissue linearity, which holds in the low strain rates of the freehand elastography, we introduced constraints on the variations of the strain field with time. We then proposed ElastMI, an EM algorithm that exploits these constraints for estimating displacement fields using

three images. The algorithm involves solving sparse linear systems, and therefore runs in real-time. The low variance motion field that we compute by exploiting this new prior can be used in numerous applications in ultrasound imaging; we used it here to generate strain images.

Acknowledgements. H. Rivaz is supported by the Predoctoral Traineeship Award from the CDMRP and by the Advanced Simulation Fellowship from the Link Foundation. We thank P. Foroughi, I. Fleming and M. van Vledder for valuable discussions and anonymous reviewers for constructive feedback.

References

1. Greenleaf, J., Fatemi, M., Insana, M.: Selected methods for imaging elastic properties of biological tissues. Annu. Rev. Biomed. Eng. 5, 57–78 (2003)
2. Hall, T., Zhu, Y., Spalding, C.: In vivo real-time freehand palpation imaging. Ultrasound Med. Biol. 29, 427–435 (2003)
3. Zahiri, R., Salcudean, S.: Motion estimation in ultrasound images using time domain cross correlation with prior estimates. IEEE TBME 53, 1990–2000 (2006)
4. Rivaz, H., Boctor, E., Foroughi, P., Fichtinger, G., Hager, G.: Ultrasound elastography: a dynamic programming approach. IEEE Trans. Med. Imag. 27, 1373–1377 (2008)
5. Rivaz, H., Fleming, I., Assumpcao, L., Fichtinger, G., Hamper, U., Choti, M.A., Hager, G.D., Boctor, E.M.: Ablation monitoring with elastography: 2D *in-vivo* and 3D *ex-vivo* studies. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 458–466. Springer, Heidelberg (2008)
6. Rivaz, H., Foroughi, P., Fleming, I., Zellars, R., Boctor, E., Hager, G.: Tracked regularized ultrasound elastography for targeting breast radiotherapy. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5761, pp. 507–515. Springer, Heidelberg (2009)
7. Lindop, J., Treece, G., Gee, A., Prager, R.: 3D elastography using freehand ultrasound. Ultrasound Med. Biol. 32, 529–545 (2006)
8. Jiang, J., Hall, T., Sommer, A.: A novel image formation method for ultrasonic strain imaging. Ultrasound Med. Biol. 53, 1088–1102 (2007)
9. Lindop, J., Treece, G., Gee, A., Prager, R.: An intelligent interface for freehand strain imaging. Ultrasound Med. and Biol. 34, 1117–1128 (2008)
10. Chen, L., Housden, R., Treece, G., Gee, A., Prager, R.: A data weighting scheme for quasistatic ultrasound elasticity imaging. Univ. of Cambridge TR651 (2010)
11. Brown, M., Burschka, D., Hager, G.: Advances in computational stereo. IEEE Trans. Pattern Anal. Mach. Intell. 25, 993–1008 (2003)
12. Moradi, M., Mousavi, P., Abolmaesui, P.: Tissue typing using ultrasound rf time series: Experiments with animal tissue samples. Med. Phys. 37, 1401–1413 (2010)
13. Li, L., Cui, Y., Kadour, M., Noble, A.: Elasticity reconstruction from displacement and confidence measures of a multi-compressed ultrasound rf sequence. IEEE Trans. Ultrason. Ferroelectr. Freq. Control 55, 319–326 (2008)
14. Catheline, S., et al.: Measurement of viscoelastic properties of homogeneous soft solid using transient elastography. J. Acoust. Soc. Amer. 116, 3734–3741 (2004)
15. Eskandari, H., Salcudean, S., Rohling, R.: Viscoelastic parameter estimation based on spectral analysis. IEEE T. Ultrason. Ferro. Freq. Control 55, 1611–1625 (2008)
16. Rivaz, H., Boctor, E.M., Choti, M.A., Hager, G.D.: Real-time regularized ultrasound elastography. IEEE Trans. Med. Imaging 30, 928–945 (2011)

Monitoring of Tissue Ablation Using Time Series of Ultrasound RF Data

Farhad Imani[1], Mark Z. Wu[1,⋆], Andras Lasso[1], Everett C. Burdette[3], Mohammad Daoud[2], Gabor Fitchinger[1], Purang Abolmaesumi[2], and Parvin Mousavi[1]

[1] Queen's University, Canada
[2] University of British Columbia, Canada
[3] Acoustic MedSystems Incorporation, USA

Abstract. *Purpose:* This paper is the first report on the monitoring of tissue ablation using ultrasound RF echo time series. *Methods:* We calcuate frequency and time domain features of time series of RF echoes from stationary tissue and transducer, and correlate them with ablated and non-ablated tissue properties. *Results:* We combine these features in a nonlinear classification framework and demonstrate up to 99% classification accuracy in distinguishing ablated and non-ablated regions of tissue, in areas as small as $12mm^2$ in size. We also demonstrate significant improvement of ablated tissue classification using RF time series compared to the conventional approach of using single RF scan lines. *Conclusions:* The results of this study suggest RF echo time series as a promising approach for monitoring ablation, and capturing the changes in the tissue microstructure as a result of heat-induced necrosis.

Keywords: tissue ablation, ablation monitoring, RF timer series.

1 Introduction

Primary and metastatic cancers of the liver cause significant distress, drastically decrease life expectancy for patients worldwide, and have an increasing frequency in North America [1]. Surgical resection, and in some cases liver transplantation, are amongst potential interventions; however, only a small proportion of patients are candidates for these treatments [2]. As a result, there has been increasing interest in ablation, including chemical ablation, cryoablation, and heat ablation, as an alternative mode of therapy [11]. Ablation therapy can be focused to specific tumor locations, thus preserving most of the surrounding tissue. In addition to metastatic disease, multiple primary tumors may not be resectable. As a result, ablation can be utilized as a minimally invasive procedure, increasing the number of patients that can receive therapy.

In heat ablation therapy, the ablation applicator is placed within the target tissue and heat is transferred to the surrounding tissue causing necrosis of the

⋆ Farhad Imani and Mark Z. Wu contributed equally to this work.

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 379–386, 2011.

target area. Microwave, laser, and Radio Frequency (RF) signals [11] have all been used as sources for heat ablation, with RF ablation being the method of choice for many clinicians. Despite advances in ablation therapy, real-time monitoring of the extent of ablation is a major limitation. The need to monitor the ablation zone is the key step to avoiding necrosis of healthy tissue and complete targeting of primary or metastatic tumors. Non-invasive monitoring techniques using magnetic resonance imaging (MRI), computed tomography (CT) and ultrasound imaging have been used for this purpose. A series of T1 and T2 MRI measurements were found to characterize changes in a variety of ex vivo tissues as a result of thermal coagulation [3], while computed tomography (CT) has been shown to effectively monitor RF ablation zones in porcine liver [10]. Despite these successes, MRI and CT-based ablation monitoring suffer several drawbacks, including requiring specialized equipment for compatibility, limitations for real-time application, and universal availability.

Ultrasound-based ablation monitoring has garnered interest due to its wide availability and relative low cost. Conventional B-mode imaging does not clearly indicate the ablation zone, and in some cases can overestimate the extent of ablation [9]. Ultrasound elastography has emerged as a technique to augment conventional images for monitoring the ablation zone [9]. Following a mechanical stimulus, various elastic properties of the tissue have been calculated from ultrasound images to identify stiffer ablated lesions.

Another mechanism to augment conventional ultrasound is to utilize raw RF signals; these signals contain tissue typing information and have been previously used for monitoring changes in frequency-dependent attenuation to determine temperature [4]. RF time series signals have also been effectively used for tissue typing at both high- and clinical-frequencies [8]. RF time series was acquired from stationary tissue and transducer over a few seconds. Frequency domain features and fractal dimension of these signals were used to distinguish between various tissue types, including healthy and cancerous prostate tissue [8].

In this paper, we aim to classify ablated and non-ablated regions of interest (ROI) in animal tissue following thermal ablation using time and frequency domain features of calibrated RF time series data. We show that the classification accuracies obtained by these features significantly outperform those achieved by spectral analysis of individual RF scan lines. The remainder of this paper is organized as follows: Section 2 presents the ablation data collection setup, feature extraction, and tissue classification approaches. Section 3 presents the results and discussion followed by concluding remarks and a summary in Section 4.

2 Methods

To differentiate between ablated and non-ablated tissue using RF time series, we acquired data from homogeneous, fresh chicken breast tissue [12]. Three experiments were performed to collect data for monitoring the ablation process. An ultrasound intersitial thermal therapy (USITT) applicator (Acoustic MedSystems, Champaign, IL) is used for tissue coagulation. In each experiment, a piece

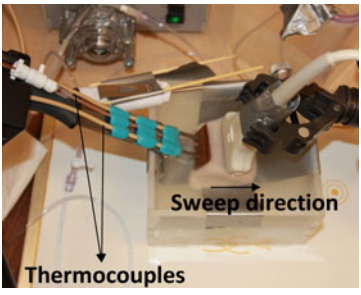


Fig. 1. Tissue ablation and data acquisition setup

of lean chicken breast tissue is placed in a warm water bath ($37°C$) with generous water flow surrounding the tissue as depicted in Fig. 1. The water bath is then placed on a warming plate to ensure the temperature stays constant. The ablator is attached to two multi-sensor type T thermocouples (Physitemp Instruments, Clifton, NJ), and inserted in the tissue. The thermocouples are used to monitor the temperature of the tissue during ablation; each thermocouple contains three sensors located at the tip, and 0.5 cm and 1 cm from the tip. A water pump is used to cool the ablator with a water flow rate of 25 ml/min. The ablator is configured to transmit RF waves with a power of 21.5 W. Ablation is initiated while the thermal dose to the tissue increases. The experiment continues until at least one sensor displays a reading of $45°C$.

RF time series data is acquired at multiple points during the experiments: prior to ablation, at the end of ablation, and after a cooling period following ablation. Sample B-mode images prior to and at the end of ablation are shown in Figure 2. As seen, the ablation zone is not obvious in the B-mode images. Ultrasound data acquisition is performed using an Ultrasonix RP scanner (Ultrasonix, Richmond, BC) with a L14-5/38 linear array transducer (Ultrasonix, Richmod, BC) operating at a central frequency of 6.6 MHz. The maximum depth of imaging is 4.5 cm while the focal point is set to 2 cm. For RF time series acquisition, 178 frames of RF data are acquired at a frame rate of 45 frames/sec. Each RF frame has 256 scan lines in the lateral direction, each with 1080 samples in the axial direction. The coupling medium for ultrasound imaging is water.

For calibration purposes, RF data is also collected from a thick piece of glass placed in the water bath at the focal point of the transducer, with imaging parameters identical to the ablation experiments.

2.1 ROI Selection and Feature Extraction

To characterize ablated and non-ablated tissue, frequency and time domain features of calibrated RF time series are calculated over ROIs. RF time series from cooled tissue at the end of ablation is used for our analysis. After aligning the orientation of the ultrasound and histology images, the tip of the ablator is used as a landmark for selection of the ROIs. Tissue characterization is compared with that of spectral features of a single RF frame over the same ROIs.

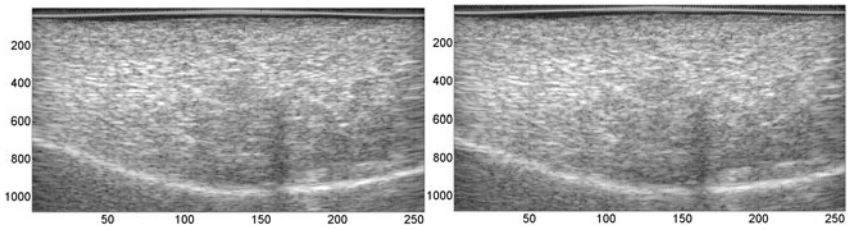


Fig. 2. B-mode images of the tissue (a) prior to, and (b) at the end of ablation. Note that the ablation zone is not visible on the ultrasound image after ablation.

A set of 120 equally-sized ROIs from ablated and non-ablated regions of the tissue are selected in each experiment, resulting in a total of 360 ROIs. The size of each ROI is 54 samples of RF from 4 scan lines that translate to a physical dimension of $5.93mm \times 2.08mm$. The ratio of the number of selected ablated to non-ablated ROIs is 1:1 (i.e. a total of 180 ablated and 180 non-ablated ROIs from three experiments). Following ablation, cooling, and data acquisition, the tissue specimen is cut approximately at the same plane as the ultrasound image acquisition, as seen in Fig. 3. The ablated zone of the tissue is clearly depicted. A crude manual registration approach is used to align ultrasound data with the image of the tissue slice at the same plane, and ROI selections are made from the ablated and non-ablated tissue accordingly.

RF time series data is calibrated for transducer effects, and used to extract features that are only related to the tissue properties. Calibration has been previously used in the literature prior to extracting spectral properties of RF ultrasound data to deconvolve the point spread function of the ultrasound imaging system [5]. Therefore, the deconvolved ultrasound RF signal is dominated by the response function of the tissue. Similarly, we use the RF data acquired from a piece of glass with high reflection. For each time frame of RF time series, its deconvolution with the glass RF signal is calculated and used as calibrated data. To compute this, the Fourier transform of a frame is divided by the Fourier transform of the glass RF signal; the calibrated RF data is computed by taking the inverse Fourier transform of this division. The time series data is then passed through a hamming window to remove the effect of artifacts in the frequency domain due to sharp edges of the rectangular windowing.

The calculated features of the calibrated RF time series are: i) the integral of the power spectrum of the first to the fourth quarter of its frequency range (Features 1-4); ii) the intercept and slope of a line fitted to the average power spectrum of the calibrated RF time series in each ROI (Features 5-6, respectively); iii) the trend of the power spectrum in different frequency bands captured by the slope of the power spectrum over the first to the fourth quarter of the frequency range (Features 7-10); and iv) the fractal dimension of the calibrated RF time series calculated using Higuchi's algorithm (Feature 11) [8].

Spectral analysis of single RF frames has been applied to tissue typing during the past decades [5]. We evaluate the performance of RF time series for characterizing ablated tissue by comparing it to the performance of the spectral features of a single RF frame from cooled tissue at the end of ablation. To extract spectral features, each scan line of an RF frame is passed through a Hamming window, followed by calibration where the power spectrum of an ROI is divided by the power spectrum of the RF data from the glass. The intercept and the slope of the regression line to the power spectrum, and the average value of the spectrum over the measurement bandwidth (midband fit) constitute the spectral features (Features 12-14). It is worth mentioning that we compensate for the effect of depth on the slope and the midband fit features using a linear attenuation coefficient similar to [5].

2.2 Classification

A support vector machine (SVM) classifier is used to classify ablated and non-ablated ROIs [6]. This classifier finds an optimal hyperplane with maximum margin separating the data into two classes. The parameters of this hyperplane are tuned by trying different kernel functions to map the training vector to higher dimensions. Among different kernel functions, the radial basis function is easier to initialize and outperforms others in terms of accuracy.

We follow a five-fold cross validation strategy. In other words, we randomly partition the ROIs into an 80%-20% ratio. We select 80% of the ROIs as training data and the remainder for testing, and repeat this for all 5 portions of the data. The entire process of five-fold cross validation is repeated 100 times, each time with random partitioning of the ROIs into five folds, to remove the effect of bias in data division. The mean classification accuracies and standard deviations are calculated over these 500 trials.

We also perform an exhaustive search for all 11 time series and three spectral features to find the optimal feature group in each category for characterizing the ablated tissue. To choose the optimal subset of 11 features, different subsets should be examined. We try subsets of 2-10 features of the time series automatically. The optimal search is done similarly for subsets of size two and three for the spectral features. The optimal classification accuracy and the best feature subset are found for each classification.

3 Results and Discussion

Ablated Tissue Classification: Two categories of features including RF time series and conventional single RF frame features are extracted from the ROIs. We performed a separate exhaustive search on each category and each experiment to identify optimal features for characterizing ablated tissue. In addition, we pooled all ROIs from the three experiments and searched for optimal features of RF time series and single RF frame.

Mean classification accuracies and standard deviations from 500 cross validation runs for individual experiments and pooled data are presented in Table 1.

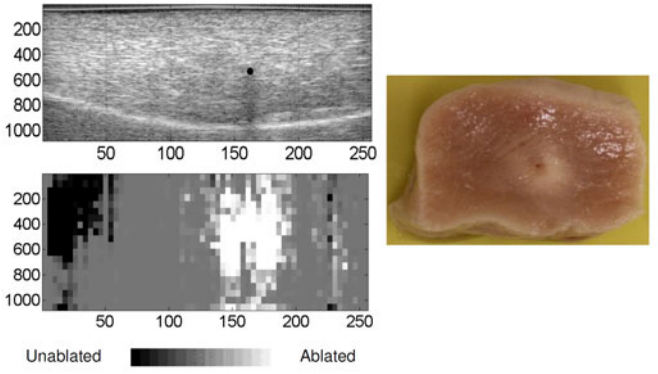


Fig. 3. An example of classifying entire image. Note the black circle in the B mode image indicating the location of the ablator, and the shadow underneath.

As seen in this table, almost perfect characterization of ablated and non-ablated tissue regions are made using only three to six RF time series features. The integral of the power spectrum of RF time series in the fourth quarter of the frequency band (Feature 4) is the single dominant feature for ablated tissue characterization in all three experiments and the pooled ROIs (rows 1-4 of the table). In addition, the intercept and the slope of the line fitted to the entire power spectrum, and the slope of the line fitted to the first frequency quarter of the power spectrum of the RF time series are the main features contributing to tissue typing. We calculated the Receiver Operating Characteristic (ROC) of tissue classification for pooled data from all experiments using the optimal time series features. The area under ROC curve obtained from RF time series features is close to one. Note that we tested RF time series data both immediately after ablation, and following a cool down period at the end of ablation. The results reported here are from the RF time series of cooled tissue; however results from the data at the end of ablation were very similar to those above.

In addition to classifying the ROIs of ablated and non-ablated tissue, we used the trained SVM classifier to test the accuracy of tissue characterization in the entire image. The B-mode ultrasound image, a color-map of the predicted posterior probabilities of ablation, and the tissue image at approximately the ultrasound imaging plane showing the physical extent of ablation are presented in Fig. 3. The ablation zone color-map is created by using all 11 features from ROIs in that particular experiment to train a model, which is then tested on the entire imaging plane. As seen, we are able to capture the ablated zone, using only a limited number of ROIs for training. The gray zone in the middle of the figure is mostly due to the deconvolution process involved in the calibration phase which results in discontinuity of the features between the RF scan lines.

Comparison with Tissue Characterization Using Conventional Single RF Frame: Similar to RF time series features, the combination of features from a single RF frame that results in the best ablated tissue characterization

Table 1. Mean and standard deviation of classification accuracies using optimal RF time series features, and conventional RF spectral features

Experiment	Combination of time series features (mean ± std)	Combination of spectral features (mean ± std)
Experiment 1	Features 3,4,6,7,10,11 (99.09±0.24%)	Features 12,13,14 (82.3±2.84%)
Experiment 2	Features 3,4,5 (98.01± 0.64%)	Features 12,14 (92.05± 1.69%)
Experiment 3	Features 4,5,6,7 (98.2± 0.29%)	Features 12,13,14 (72.41± 2.89%)
Pooled Experiments	Features 4,5,6,7,9 (99.29± 0.24%)	Features 12,13,14 (75.59± 1.28%)

are presented in Table 1. For most of the experiments as well as data from pooled experiments, all three spectral features are required. Comparing the columns of Table 1 shows that classification accuracies achieved using optimal time series features are above 98% for all of the experiments, whereas accuracies obtained incorporating the best spectral features vary between 72% and 92%. Furthermore, the standard deviations of the classification accuracies obtained by time series features are superior as well. When the ROIs of three experiments are mixed, the classification accuracy obtained using the combination of time series features is 99.29%. For the same ROIs, the classification accuracy computed by using spectral features is 75.59%. Finally, the area under the ROC curve for the RF time series features is close to 1 compared to 0.8 for spectral features.

Previously, it was shown that temperature rise produced by acoustic propagation may be the major contributor to the ability of time series analysis for tissue typing [7]. In other words, the speed of acoustic waves in the tissue is dependent on the tissue temperature. Ultrasound frames from RF time series can be used to estimate phase delay, hence virtual displacement. In our preliminary results, we have also observed that there is an upward trend for the phase delay of the ablated ROIs. Furthermore, the amplitude and frequency components of phase delay are distinguishable between ablated and non-ablated ROIs.

4 Conclusion

Monitoring the ablation zone is essential in ablation therapy to avoid necrosis of healthy tissue, and to completely target tumors. RF time series has shown great promise for tissue typing at both high and clinical frequencies. In this paper, we report for the first time the application of frequency and time domain features of RF time series for characterizing ablated tissue. We calibrate the RF signals of each individual frame to remove the contribution of the ultrasound imaging system. We calculate eleven features from the time series of RF signals of 360 ROIs in three animal experiments. We use an SVM classifier to characterize tissue into ablated and non-ablated categories, within a five-fold cross validation

framework. The results show classification accuracies as high as 99% are obtained using these features by analyzing ROIs with sizes of $12mm^2$. In addition, time series features significantly outperform spectral features of a conventional RF frame in terms of the classification mean accuracy and standard deviation, and the area under ROC curve. Future work will focus on expanding the current approach to real-time monitoring of the ablation zone in the entire imaging plane. Towards this goal, we extend the classification framework to the entire plane and demonstrate color-maps of the ablation zone that mimic the tissue histology closely. More work needs to be done on more accurate registration of the histology to ultrasound images, in addition to further experiments for generalization of the ablation zone beyond ROIs.

References

1. Nakakura, E.K., Choti, M.A.: Management of hepatocellular carcinoma. Oncology (Williston Park) 14(7), 1085–1102 (2000)
2. Steele, G.J.: Colorectal Cancer Metastatic to the Liver Resection. In: Cameron, J.L. (ed.) Current surgical therapy, 5th edn., pp. 283–289. St. Louis (1995)
3. Graham, S.J., Stanisz, G.J., Kecojevic, A., Bronskill, M.J., Henkelman, R.M.: Analysis of Changes in MRI Properties of Tissues After Heat Treatment. Magn. Reson. Med. 42(6), 1061–1071 (1999)
4. Ueno, S., Hashimoto, M., Fukukita, H., Yano, T.: Ultrasound Thermometry in Hyperthermia. In: Proc. of the IEEE Ultrasonics Symp., pp. 1645–1652 (1990)
5. Lizzi, F., Greenebaum, M., Feleppa, E., Elbaum, M.: Theoretical Framework for Spectrum Analysis in Ultrasonic Tissue Characterization. J. Acoust. Soc. Am. 73(4), 1366–1373 (1983)
6. A Library for SVMs, http://www.csie.ntu.edu.tw/~cjlin/libsvm/
7. Daoud, M.I., Mousavi, P., Imani, F., Rohling, R., Abolmaesumi, P.: Computer-aided Tissue Characterization Using Ultrasound-induced Thermal Effects: Analytical Formulation and In vitro Animal Study. In: SPIE, Orlando (2011)
8. Moradi, M., Abolmaesumi, P., Siemens, D.R., Sauerbrei, E.E., Boag, A., Mousavi, P.: Augmenting Detection of Prostate Cancer in Transrectal Ultrasound Images Using SVM and RF Time Series. IEEE Trans. Biomed. Eng. 56(9), 2214–2223 (2009)
9. Boctor, E., deOliveira, M., Choti, M., Ghanem, R., Taylor, R., Hager, G., Fichtinger, G.: US 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)
10. Cha, C.H., Lee, F.T., Gurney, J.M.: CT versus sonography for monitoring radiofrequency ablation in a porcine liver. Am. J. Roentgenol. 175, 705–711 (2000)
11. Lau, W.Y., Leung, T.W.T., Yu, S.C.H., Ho, S.K.W.: Percutaneous Local Ablative Therapy for Hepatocellular Carcinoma: A Review and Look Into the Future. Ann. Surg. 237(2), 171–179 (2003)
12. Zheng, X., Vaezy, S.: An Acoustic Backscatter-based Method for Localization of Lesions Induced by High-intensity Focused Ultrasound. Ultrasound in Med. and Biol. 36(4), 610–622 (2010)

Evaluation of *in vivo* Liver Tissue Characterization with Spectral RF Analysis versus Elasticity

Stéphane Audière[1,2], Elsa D. Angelini[1], Maurice Charbit[1], and Véronique Miette[2]

[1] Institut Telecom, Telecom ParisTech, CNRS LTCI, Paris, France
[2] Echosens, Research and Development Department, Paris, France

Abstract. Ultrasonic elastography, via vibration-controlled transient elastography (VCTE™), enables to assess, under *active* mechanical constraints, the elasticity of the liver, correlating with fibrosis stages. On the other hand, the same VCTE™ probe can also be used in *passive* mode, acquiring RF lines at different locations in the liver. This paper presents a thorough evaluation of passive-mode RF spectral parameters (integrated backscatter coefficient, power spectral index, effective scattering size and spectral variance), for tissue characterization on a large cohort of volunteers with various ranges of elasticity measures. Results showed that capabilities to discriminate between liver and subcutaneous fat tissues were highly variable among spectral parameters. Furthermore, it appears that no *in vivo* discrimination of liver elasticity/fibrosis stage can be performed with passive RF spectral analysis, at 3.5MHz.

Keywords: ultrasound, RF lines, liver, backscatter coefficient, scatterer size, spectral analysis, elastography.

1 Introduction

A large amount of literature exists on the potential of backscatter radiofrequency (RF) analysis to characterize tissue content. The underlying assumption in this field is that scattering properties characterize and discriminate tissues. However, in conventional ultrasonic data processing and image formation, the frequency content of the backscattered RF data is generally not exploited. The backscattered RF ultrasound signal content depends on the acoustic properties of the tissue (density and size of the reflectors), but also on the transfer function of the transducer. Both components are frequency dependent through signal attenuation and backscattering intensity in the tissue. In this study we investigated several spectral parameters to discriminate *in vivo* different types of tissues and different stages of liver fibrosis. The study was divided into two steps: first we investigated if liver fibrosis stage correlates with some acoustic parameters that are sensitive to tissue microstructures. Secondly, we investigated if subcutaneous and liver tissues could be discriminated based on acoustic parameters that are sensitive to tissue macrostructures. Such information is necessary for adequate use of the Vibration-controlled transient elastography (VCTE™) probe. Measurements were performed on a database of 181 volunteers, with a single-transducer VCTE™ probe operating at f_0 = 3.5MHz in a bandwidth 2.5-4.5MHz and with a focus

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 387–395, 2011.

at 35mm. Similar studies appear in the literature but on smaller *in vivo* databases and with histological fibrosis scores, while we used mechanical elasticity measurements.

The ultrasound RF signals were acquired for a depth range between 5 and 65 mm, to accommodate for the variability of the skin-liver capsule distance within the volunteers population. Pass-band filtering was applied in band 1-8 MHz by the frontal hardware. From individual RF lines, we extracted the following parameters: the broadband ultrasound attenuation (BUA), the backscattered coefficient (BSC), the scattering power index (SPI), the effective scatterer size (ESS) and the spectral variance (SV). Screening was also performed in active (elastography) mode, using the VCTE™ technique [1] to quantify liver elasticity by measuring the velocity of a low-frequency shear wave generated by a pulse and travelling through the liver. It has been demonstrated that stiffness is highly correlated with fibrosis stage assessed by liver biopsy [2].

2 Materials and Methods

Measured RF signals correspond to echoes of an initially transmitted pulse, which propagated up to a certain depth in the body, attenuated along the travel path and partially reflected at interfaces between tissues and by non-specular reflections in "granular" tissues such as the liver. The notion of granularity depends on the ratio of the wavelength of the emitted signal and the micro-texture of the tissue. At 3.5MHz, for an average speed of sound of 1500 m/s in human soft tissues, this corresponds to wavelength λ=430µm, which is of the order of magnitude of the liver microstructures. The spectrogram $S(f,z)$ of the RF signal at frequency f and at a distance z from the probe is generally modelled as [3-4]:

$$S(f,z) = |H(f)|^2 |R(f)|^2 e^{-2\alpha(f)z}, \tag{1}$$

where H is the Fourier Transform (FT) of the composite pulse, consisting of the transmitted pulse and the response of the electronic device. R is the FT of the backscatter component from the echoes and $\alpha(f)$ is the attenuation coefficient function. The last two terms encode the interaction of the transmitted pulse with the tissues. Within the limited bandwidth of the ultrasound transducer, experimental measures have shown that RF lines are attenuated in soft tissues proportionally to the operating frequency so that $\alpha(f) = \beta f$, where β is the BUA.

Regarding the generation of non-specular reflection in a homogeneous tissue, the backscattering spectrum is often modeled as $R(f) = f^n$, where n is the SPI that ranges from 0 to 4 in soft materials, and from 1 to 2 for most human soft tissues.

The transmitted pulse consisted of two periods of sinusoid with apodization. This signal can be adequately approximated by a modulated Gaussian pulse leading to a Gaussian spectral distribution that writes:

$$H(f) \propto e^{-\frac{(f-f_0)^2}{2\sigma^2}}, \tag{2}$$

where f_0 is the transducer central frequency and σ the pulse spectrum bandwidth.

2.1 RF Spectral Parameters Related to Tissue Characterization

Five spectral parameters were studied as illustrated in Fig. 1

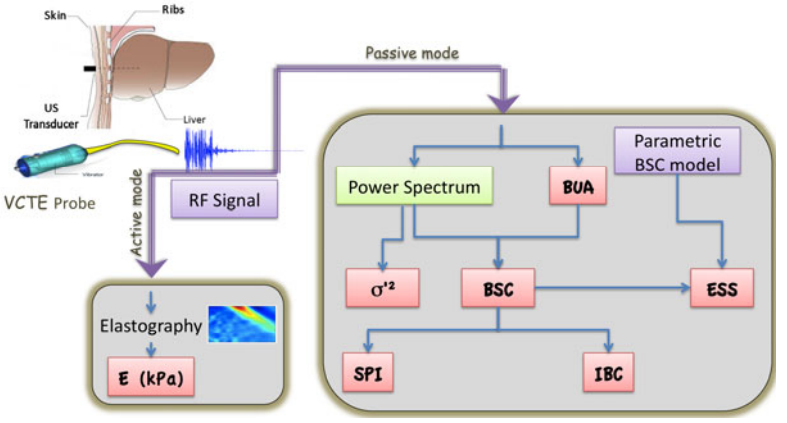


Fig. 1. Liver screening protocol (left) and flowchart of spectral RF signal analysis

We now detail the estimation of these parameters, based on various models commonly used in the literature dedicated to ultrasound-based tissue characterization.

Estimation of the BUA

Taking the log of Eq. (1), the log-spectrum writes

$$\ln(S(f,z)) = \ln\left|H(f)\right|^2 + \ln\left|R(f)\right|^2 - \alpha(f)2z\,, \tag{3}$$

Assuming that the tissue is homogeneous over a thickness of a few wavelengths, we use a short-term spectral analysis (STSA) based on the Welch's approach. The attenuation coefficient $\alpha(f)$ is then accessed via the derivative of the log-spectrum with respect to z. Indeed, for each frequency component, we have:

$$\alpha(f) = -\frac{1}{2}\frac{\partial \ln(S(f,z))}{\partial z}\,, \tag{4}$$

For the linear model, $\alpha(f) = \beta f$, the broadband ultrasound attenuation (BUA) coefficient can be estimated by least square regression on the experimental attenuation coefficient [5]. It is general practice in clinical use to reconstruct ultrasound images assuming a fixed β value of $0.5dB\ cm^{-1}MHz^{-1}$ for human soft tissues.

Estimation of the Backscatter Coefficients BSC and IBC

The backscatterer coefficient was introduced in [6] and corresponds to the relative scattering cross-section per unit solid angle and volume. Indeed to compute the BSC, echoes from the media of interest are compared to specular echoes measured from a steel plate with a reflection coefficient of R=0.9 and placed at the focal length of the transducer. This plate is used as the reference signal to learn the transducer transfer function. The BSC is then computed as:

$$BSC(f,Z) = \frac{1.45Z^2}{A_0\Delta z}W(f,Z)\,, \tag{5}$$

where A_0 is the area of the transducer aperture, Δz is the axial length used for local measures, and Z is the axial distance between the transducer and plate. $W(f,Z)$ is the ratio between the spectrograms of the backscattered RF signal and the reference signal, and given by:

$$W(f,Z)=\left(\frac{R}{2}\right)^2 e^{4(\alpha_m(f)-\alpha_0(f))(Z+\Delta z/2)}\frac{S_m(f,Z)}{S_0(f,Z)}, \quad (6)$$

where S_m is the STSA of the sample media RF signal, S_0 is the STSA of the plate RF signal, $(\alpha_m(f),\alpha_0(f))$ are the attenuation coefficients (Eq. 1) for the sample and reference media. From a series of BSC measurements at different frequencies within the bandwidth $[f_{\min}, f_{\max}]$ of the transducer, we can compute the integrated backscatter coefficient (IBC), as the frequency-average of the BSC:

$$IBC(Z)=\frac{1}{f_{\max}-f_{\min}}\int_{f_{\min}}^{f_{\max}} BSC(f,Z)df, \quad (7)$$

Estimation of the Scattering Power Index (SPI)

For independent scatterers of size smaller than the wavelength, theoretical approaches [7] lead to a BSC frequency dependence in f^4 : this is called the Rayleigh limit. In practice, it was observed that this simple model fails in soft tissues [8], leading to a generalization of the BSC frequency dependence model as: $BSC(f,Z)\propto f^n$, where n is the scattering power index. To estimate the SPI, a log-log linear regression analysis on the empirical BSC with respect to frequency within the bandwidth of the transducer is performed. The regression slope yields a tissue-specific SPI value which ranges between 1 and 4 in soft tissues, as reported in several studies [9-10].

Estimation of the Effective Scatterer Size (ESS)

Rather than relying on a macroscopic measure of the backscatterer intensity with the BSC, alternative theoretical backscattering models have been proposed [9-10]. These models provide theoretical BSC expressions as the TF of the autocorrelation functions explicitly parameterized with the effective scatterer size a, corresponding to the spatial variation of the density and compressibility within the tissue. A series of BSC curves was generated for different values of a. From the slope of the log-log linear regression, a series of *theoretical* SPI values, related to a, was derived. The a value providing the theoretical SPI closest to the empirical SPI is then selected as the scatterer size of the observed tissue. In this study we tested two autocorrelation models [4, 9], described in Table 1, to estimate the ESS.

Table 1. Spatial autocorrelation models and theoretical BSC functions

Autocorrelation model	Theoretical BSC	Parameters
Gaussian	$8R_S^2k^4a^3e^{-k^2a^2/4}/\sqrt{\pi}$	$a = ESS = [0\ \ 300]\mu m$ $R_S = R_i\sqrt{Na^3+N^2a^6}$ $R_i = reflection\ coefficient = 1$ $N = scatterer\ density$ $k = 2\pi f/c,\ f = 3.5MHz$ $c = speed\ of\ sound = 1500m.s^{-1}$
Exponential	$64R_S^2k^4a^2/\pi(1+(ka)^2)^2$	

Spectral Variance σ'^2

As in [11], the spectrogram of the received echoes $S(f,z)$ can be approximate as a Gaussian pulse, around the central frequency f_C following:

$$S(f,z) \propto e^{\frac{(f-f_c)^2}{2\sigma'^2}}, \tag{8}$$

where σ' is given by:

$$\sigma'^2 = \sigma^2 \frac{f_0^2}{f_0^2 + n\sigma^2}, \tag{9}$$

According to this model, soft tissues with different acoustic properties (characterized by *n*) should generate echoes with different spectrum bandwidth. Computation of this parameter on *in vivo* RF signals is detailed in [12].

The literature on RF based tissue characterization, exploiting all the previous models, contains very few *in vivo* studies, and only performed on small cohorts of volunteers. We exploited a large cohort of volunteers, to evaluate the models, and investigate the potential correlation between elasticity measures and spectral RF parameters for a large range of values.

3 Results and Discussion

3.1 Ultrasound RF Scanning Setup and Procedure

The VCTE™ device used in this study was composed of a probe containing a low-frequency vibrator, an ultrasonic transducer operating, at 3.5 MHz, a dedicated electronic system and a control unit (50 MHz sampling frequency with a 14-bit resolution). A single element probe was used both as an emitter and a receiver. For each volunteer, 400 ultrasound RF lines were acquired, along a fixed scan line, at a pulse repetition frequency of 20 Hz during 20 s.

A database was acquired with the VCTE™ probe, operating in passive and elastography modes [1], on 181 volunteers. Elastography mode was used to measure the elasticity of the liver. The ultrasound RF lines were acquired for a depth range between 5 and 65mm, to accommodate for the variability of the skin-liver capsule distance within the population.

A phantom was also used to validate our computational framework for the experimental BSC, SPI and ESS computations. It was made of a mixture of cohesive copolymer, white mineral oil, and silica powder as acoustic scatterers [13]. Particle diameters ranged between 35–70 μm. The diameter of these particles, smaller by an order of ten than the wavelength of the ultrasound pulse (λ=430 μm) provides Rayleigh scattering conditions (i.e. BSC proportional to f^4).

Spectral parameters were estimated on individual RF lines, using STSA with 75% overlapping windows of length 12λ, along the whole depth range. Spatial parameter values can be displayed as parametric images for tissue segmentation, as [14], or concatenated into their median value for quantitative characterization of tissue types (such as liver tissues with different fibrosis stages).

3.2 Quantitative Liver Fibrosis Stage Characterization via Spectral Analysis

Phantom Study

Average elasticity of the phantom was measured, using the VCTE™ elastography probe, as E=6.0±0.05 kPa over 10 measurements. Median values over the 400 RF lines provided attenuation and BSC measures illustrated in Fig. 2. We can observe on these results that the phantom media behaved as expected, with attenuation and BSC coefficients varying linearly, in frequency and in log-frequency respectively, within the bandwidth of the transducer. Linear interpolations, illustrated in Fig. 2, lead to empirical estimations of the BUA=1.14±0.03 dB/cm/MHz and SPI= 3.97±0.15. The other spectral parameters provided the following measures: IBC= -68.14±0.89dB, ESS =19.85±9.24 µm with the exponential model, ESS=55.9±26.16 µm with the Gaussian model and σ'^2=1.36±0.05. These results were in agreement with what was expected, with a SPI close to 4 and ESS size close to the phantom, specifications using the Gaussian autocorrelation model.

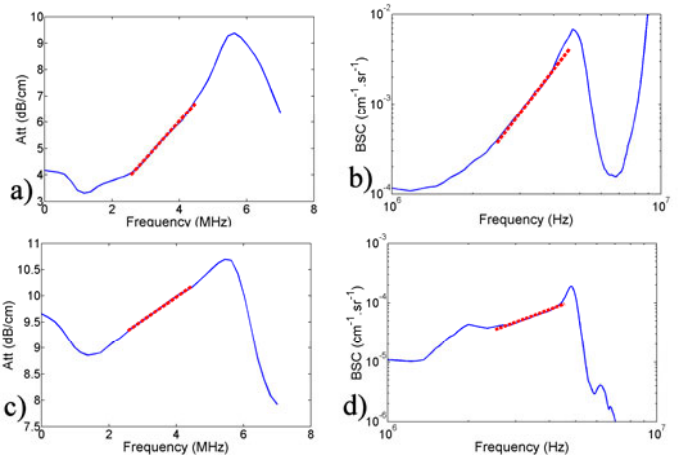


Fig. 2. Experimental measures on a phantom(a-b) and liver (c-d): (a-c) Attenuation and BUA measurements. (b-d) BSC and SPI measurements.

***In vivo* Study**

For *in vivo* RF signals, we report, in Table 2, median values of the spectral parameters computed on the 400 RF lines. We have considered three classes of liver fibrosis [2]. In Table 2, *N* denotes the number of volunteers in each class.

Table 2. Average-median spectral parameters on *in vivo* liver tissue scanning

Fibrosis stage	***N***	**E** (kPa)	**BUA** (dB/cm/MHz)	**SPI** (f^n)	**ESS** (µm)	**IBC** (dB)	**σ'** (MHz)
Non to mild	105	4.7±0.8	0.41±0.23	1.68±0.68	220.5±36.8	-58.5±6.8	1.03±0.03
Significant to extensive	66	7.2±1.3	0.37±0.29	1.73±0.88	218.0±44.9	-59.9±8.3	1.05±0.03
Cirrhosis	10	21.0±9.2	0.38±0.33	1.81±0.94	214.3±51.2	-58.8±9.7	1.06±0.04

BUA and BSC are reported Fig. 2. We observed that linear fits on the attenuation and the BSC curves was suitable for *in vivo* RF measures, within the bandwidth of the transducer. None of the average parameter measures was able to discriminate between the three classes of liver fibrosis stages. This result is in agreement with the *in vivo* study of [15] but which was performed on a cohort of only 17 volunteers.

3.3 Spectral Parametric Images for Tissue Segmentation

Parametric images were generated for all the spectral parameters. As illustrated in Fig. 3 for three of the parameters, these images showed high spatial variations of the parameter values and poor homogeneity within the three tissue layers: subcutaneous fat, muscle/fat layer and liver.

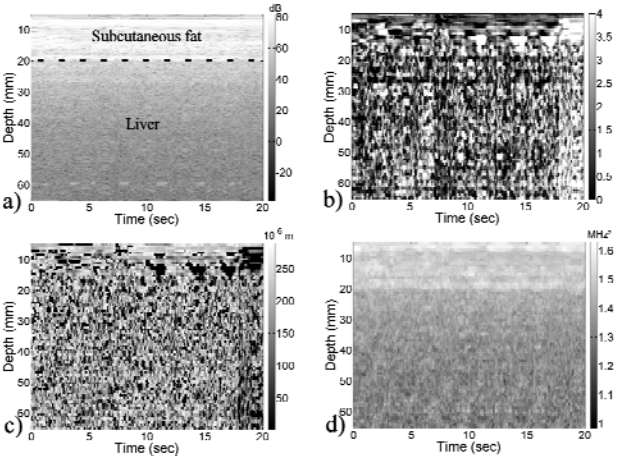


Fig. 3. Image of RF line envelopes (a) and parametric maps of SPI (b), ESS (c) and σ'^2 (d) spectral parameters

Segmentation of these layers might be worthwhile to investigate via three approaches: (1) thresholding of the RF envelope, which is limited by a high noise level and uncalibrated variations on the RF envelope intensity, (2) thresholding of the spectral variance, confirming the results reported in [12], (3) exploitation of the SPI converging to zero at specular interfaces between the three layers, as clearly observed in Fig. 3. Regarding the ESS parameter, it is not discriminating for small SPI values and can only reflect macro granularity of the tissues, rather than characterizes interfaces.

4 Conclusion

This paper has presented an *in vivo* study of spectral parametric analysis of RF ultrasound signals for tissue characterization. Results were in agreement with the literature and showed that *in vivo* spectral tissue characterization remains challenging. While

spectral parametric analysis provides promising capabilities for tissue segmentation, fine classification of tissue fibrosis stage seems too difficult. Maybe the experimental conditions were not appropriate for accurate quantification of scattering properties. Indeed, one important limitation is that all these parameters (except the SV) rely entirely on an accurate estimate of the BUA that requires a homogeneous tissue, which is not the case in the fat layer. In the majority of papers, the problem was acknowledged too challenging and an average value of 0.5dB/cm/MHz is used.

From our results, we conclude that no *in vivo* discrimination of liver elasticity/fibrosis stage can be performed with passive RF spectral analysis, at 3.5MHz and a wavelength of 430µm. We infer from this that either liver tissue homogeneity must be studied at a shorter wavelength, which is not applicable *in vivo* due to attenuation effects, or we need to look at the tissue structures at a larger scale, for example with the mean scatterer spacing [16], which is a non-parametric approach requiring higher frequencies in the echo pulses.

References

[1] Sandrin, L., et al.: Transient elastography: A new noninvasive method for assessment of hepatic fibrosis. Ultrasound in Medicine and Biology 29, 1705–1713 (2003)
[2] Castera, L., et al.: Prospective comparison of transient elastography, Fibrotest, APRI, and liver biopsy for the assessment of fibrosis in chronic hepatitis C. Gastroenterology 128, 343–350 (2005)
[3] Oosterveld, B.J., et al.: Ultrasound attenuation and texture analysis of diffuse liver disease: methods and preliminary results. Physics in Medicine and Biology 36, 1039–1064 (1991)
[4] Romijn, R.L., et al.: Estimation of scatterer size from backscattered ultrasound: a simulation study. IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control 36, 593–606 (1989)
[5] Kuc, R., Schwartz, M.: Estimating the Acoustic Attenuation Coefficient Slope for Liver from Reflected Ultrasound Signals. IEEE Transactions on Sonics and Ultrasonics 26, 353–361 (1979)
[6] Insana, M.F., et al.: Describing small-scale structure in random media using pulse-echo ultrasound. Journal of Acoustical Society of America 87, 179–192 (1990)
[7] Wagner, R.F., et al.: Statistical properties of radio-frequency and envelope-detected signals with applications to medical ultrasound. J. Opt. Soc. Am. A 4, 910–922 (1987)
[8] Lizzi, F.L., et al.: Relationship of ultrasonic spectral parameters to features of tissue microscructure. IEEE Transactions on Ultrasononics, Ferroelectrics and Frequency Control 33, 319–333 (1986)
[9] Ueda, M., Ozawa, Y.: Spectral analysis of echoes for backscattering coefficient measurement. The Journal of the Acoustical Society of America 77, 38–47 (1985)
[10] Nicholas, D., et al.: Evaluation of backscattering coefficients for excised human tissues: Principles and techniques. Ultrasound in Medicine & Biology 8, 7–15 (1982)
[11] Treece, G., et al.: Ultrasound attenuation measurement in the presence of scatterer variation for reduction of shadowing and enhancement. IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control 52, 2346–2360 (2005)

[12] Audière, S., Charbit, M., Angelini, E.D., Oudry, J., Sandrin, L.: Measurement of the skin-liver capsule distance on ultrasound RF data for 1D transient elastography. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6362, pp. 34–41. Springer, Heidelberg (2010)

[13] Oudry, J., et al.: Copolymer-in-oil phantom materials for elastography. Ultrasound Med. Biol. 35, 1185–1197 (2009)

[14] Insana, M.F., Hall, T.J.: Parametric ultrasound imaging from backscatter coefficient measurements: Image formation and interpretation. Ultrasonic Imaging 12, 245–267 (1990)

[15] Wear, K.A., et al.: Measurements of ultrasonic backscatter coefficients in human liver and kidney in vivo. The Journal of the Acoustical Society of America 98, 1852–1857 (1995)

[16] Machado, C.B., et al.: Characterization of in vitro healthy and pathological human liver tissue periodicity using backscattered ultrasound signals. Ultrasound in Medicine & Biology 32, 649–657 (2006)

Towards Extra-Luminal Blood Detection from Intravascular Ultrasound Radio Frequency Data

E. Gerardo Mendizabal-Ruiz[1], George Biros[2], and Ioannis A. Kakadiaris[1]

[1] Computational Biomedicine Lab, Departments of Computer Science, Electrical and Computer Engineering, and Biomedical Engineering, University of Houston, Houston, TX
[2] Department of Biomedical Engineering and School of Computational Science and Engineering, Georgia Institute of Technology, Atlanta, GA

Abstract. Recent evidence has suggested that the presence and proliferation of vasa vasorum (VV) in the plaque is correlated to an increase in plaque inflammation and destabilization, leading to acute coronary events (e.g., heart attacks). Therefore, the detection and quantification of VV in plaque (i.e., extra luminal blood perfusion) is an important problem since it may enable the development of an index of plaque vulnerability. In this paper, we explore the feasibility of a method that employs a physics-based model of the scattered intravascular ultrasound (IVUS) radio frequency signal for the detection of blood. We evaluate our method using synthetic data and validate it using six 40 MHz pullback sequences acquired with three different IVUS systems from different arteries of rabbits and swines. Our experimental results are very promising and indicate the feasibility of our method for the computation of a feature that leads to automatic extra-luminal blood detection which may be an indication of plaque inflammation.

1 Introduction

Atherosclerosis is characterized by the formation and accumulation of plaque in the walls of the arteries which results in the hardening and thickening of the arteries [8,16]. Coronary events such as heart attacks are the result of inflammation or thrombotic complications of the plaque. Vasa vasorum (VV) is a network of microvessels that penetrate and nourish the wall of the vessel [3]. Recent evidence has suggested that the presence and proliferation (i.e., increase in density) of VV in the plaque is correlated with an increase in plaque inflammation and the processes which lead to its destabilization [7]. Based on this evidence, it is believed that the detection and measurement of VV in plaque and the detection of leakage of blood within the plaques can enable the development of an index of plaque vulnerability.

Intravascular ultrasound (IVUS) is a catheter-based medical imaging modality that is capable of providing cross-sectional images of the interior of blood vessels and is currently the gold-standard technique for assessing the morphology of blood vessels and atherosclerotic plaques *in-vivo*. The IVUS catheter consists of a miniaturized ultrasound transducer which transmits ultrasound pulses and receives its acoustic radio frequency (RF) echo signals (i.e., A-line) at a discrete set of angles. The gray-scale B-mode IVUS images are the result of postprocessing (i.e., envelope detection, compression, compensation, scaling, and geometrical transformation) of these A-line signals.

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 396–403, 2011.

While there have been several efforts to automatically extract and analyze the information from by IVUS data, it has not been until recent years that research has started to focus on the analysis of the RF ultrasound signals instead of the IVUS B-mode images as they are more reliable since they are not affected by any processing or transformations. Nair *et al.* [10,11] proposed a method for plaque characterization, known as "virtual histology" (IVUS-VH), that is based on the use of features extracted from the signal's power spectrum. Kawasaki *et al.* [5] proposed a tissue classification method using features computed with from integrated backscatter of the RF signal. The feasibility of using wavelet analysis for plaque characterization using the RF amplitude was studied by Katouzian *et al.* [4] and Roodaki *et al.* [15]. Recently, Ciompi *et al.* [1] presented a method for plaque characterization that enhances *in-vitro* training sets by including examples from *in-vivo* coronary plaques using a floating forward feature selection method. Korga *et al.* [6] proposed a method for plaque characterization using fractal analysis-based features of the IVUS RF signal and a k-nearest neighbor classifier. O'Malley *et al.* [12] presented a study of the feasibility of blood characterization using IVUS data by employing features based on frequency-domain measures of the high-frequency signal. A common limitation of most of these methods is that the features that characterize the tissues of interest do not consider the effects of the interactions of the sound waves with the tissues. These effects (e.g., radial attenuation and attenuation due to the medium) determine the characteristics of the RF signal along the time axis. Therefore, the validity of a set of features may not be the same for the same type of tissues at different distances from the transducer. To overcome these limitations, our group has previously presented a method for the segmentation of the lumen in IVUS data using the RF signal and a physics-based model of the received IVUS RF signal [9]. In that work, the lumen/wall interface for each transducer angle was detected by solving an inverse problem. In this paper, we explore the feasibility of a new method for the detection of blood from IVUS using a similar approach with the following differences: (i) the new method is based on the comparison of the root mean square (RMS) power of the RF IVUS signal instead of the raw B-mode data, (ii) we perform the detection of several interfaces simultaneously instead of only the lumen and wall interface, (iii) the proposed method includes a regularization term to increase stability, and (iv) the proposed method is considerably faster since the problem can be formulated as a banded linear system which can be solved very efficiently. Specifically, our contributions are: (i) an efficient method for extracting a physics-based feature for blood detection from IVUS RF data, and (ii) a method for generating pseudo-colored B-mode images based on this feature. The rest of the paper is organized as follows: Section 2 presents the methods, Section 3 presents the results obtained, Section 4 presents our discussion, and Section 5 presents our conclusions and future work.

2 Methods

2.1 Scattering Model

When an incident sound wave interacts with an object, a fraction of its power will be reflected and a fraction will be absorbed by the object. When the wavelength of the incident wave is smaller in comparison to the size of the object, the wave is reflected in all

directions (i.e., scattering). The power scattered by each object in the direction opposite to the direction of the incident wave depends on the differential backscattering cross section (DBC), which can be considered as a measurement of the effective (acoustic) area of the object [14]. The collective interaction of all the scatterers can be modeled using the Born approximation [2] which implies that the scattered echoes are weak in comparison to the incident signal, and therefore it is possible to use the principle of superposition to represent the total scattered wave as a sum of the individual reflections of each point scatterer. By considering that the wavelength of the IVUS impulse signal is large in comparison to the structures in the vessel, we can model the received IVUS RF signal by representing the structures in the vessel as a finite set of scatterers with an associated DBC coefficient. Consider a set of N point scatterers, $\Phi = \{\phi_1, \phi_2, ..., \phi_N\}$, where each scatterer $\phi_i = \{\theta_i, r_i, \tau_i\}$ is characterized by its angular position θ, its radial distance from the transducer r, and its DBC τ. An A-line signal can be modeled by computing the interaction of the impulse wave with the set of M scatterers inside an angular window $\{\Phi_\theta : (\theta - \triangle\theta) \leq \theta \leq (\theta + \triangle\theta)\}$, where $2\triangle\theta$ is the angular divergence of the ultrasound beam. The received RF signal for the transducer angular position θ can be modeled as:

$$\hat{S}_\theta(t) = \sum_{i=1}^{M} \tau_i \frac{e^{(-\mu r_i)}}{r_i} e^{\left(\frac{-(t-\frac{r_i}{c})^2}{2\sigma^2}\right)} \sin\left(\omega\left(t - \frac{r_i}{c}\right)\right), \tag{1}$$

were, $\omega = 2\pi f$ is the angular velocity of the impulse wave of frequency f, c is the speed of sound, and σ is the standard deviation of a Gaussian function that is used to approximate the envelope of the impulse function [18].

2.2 Characterization of Blood

Our hypothesis is that we can estimate the DBC value for small partitions of the IVUS RF signal and use these values to characterize and detect blood. Let $S_\theta(t)$ and $\hat{S}_\theta(t)$ be the received and modeled RF signals, respectively, for the transducer angular position θ (i.e., A-line). These signals are divided into N_P non-overlapping partitions of the same size $\triangle P = \beta_p - \alpha_p \ \forall p \in \{1, 2, ..., N_p\}$. By assuming that the signal contained on each partition $P_{\theta,p}$ is generated by a unique type of tissue, it is considered that all the scatterers that generate that signal have the same DBC $\tau_i = \tau_{\theta,p}, \forall i : \alpha_p \leq r_i < \beta_p$. To estimate the DBC value that generates the signal in each partition we propose to compute the value $\tau_{\theta,p}$ such that the quadratic error between the RMS power of the real signal $R_{\theta,p}$ and the modeled signal $\hat{R}_{\theta,p}$ for the partition $P_{\theta,p}$ is minimal. However, the characteristics of the RF signals depend on the spatial position arrangement of the scatterers, which is unknown. Similar with our previous work [9], we employ the Monte-Carlo approach on which N_s samplings of random scatterers' positions with a given density D (i.e., number of scatterers per mm^2) are used to estimate the DBC $\tau_{\theta,p}$ such that the quadratic error between the RMS power of the real and each of the sampling modeled signals $\hat{R}^s_{\theta,p}$ is minimum. Additionally, in this work we introduce a regularization term that embodies our assumptions about the variability in the DBC values of the neighboring partitions G_δ across the angular direction. Here, δ refers to

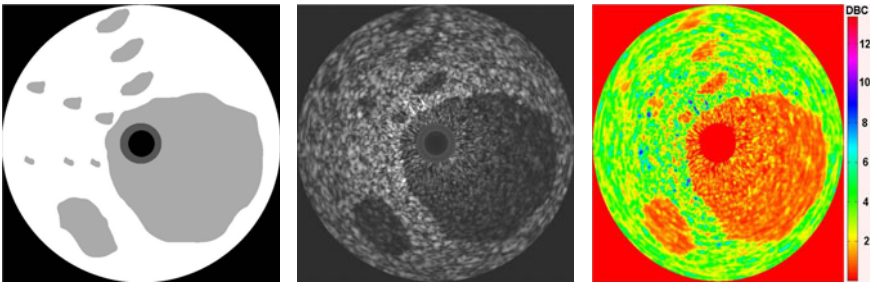
the cardinality of the neighbors and $N_n = 2\delta$ is the number of neighbors. For each partition, the DBC can be computed as:

$$\arg\min_{\tau_{\theta,p}} \sum_{s=1}^{N_s} \left(\hat{R}^s_{\theta,p}(\tau) - R_{\theta,p}\right)^2 + \beta \sum_{j \in G_\delta} \left(\tau_{\theta,p} - \tau_{j,p}\right)^2, \tag{2}$$

where β is a parameter that controls the contribution of the regularization term. The DBC for all the partitions can be efficiently computed simultaneously by solving a banded-matrix linear system.

3 Results

Synthetic data: To verify the capability of our method in recovering the DBC of the tissues of interest using the IVUS RF data, we performed experiments using synthetic IVUS RF data that were created using parameters obtained from the literature and using a mask which determine the blood and non-blood regions. The DBC values were $\tau_b = 1$ mm^2 and $\tau_w = 2$ mm^2 for blood and non-blood, respectively, while the scatterer densities were $D_b = 100$ scatterers/mm^2 and $D_w = 150$ scatterers/mm^{-2} for blood and non-blood, respectively. Since the exact values of DBC are known, it is possible to assess the sensitivity of our method with respect to the parameters such as the size of partition, number of samplings, value of the regularization term parameter, and the cardinality of the neighbors (Fig. 2). The mask used for creating the synthetic data, its corresponding IVUS B-mode reconstruction, and the recovered DBC values using 100 samplings, $\triangle P = 0.05$ *mm*, $\beta = 1$, and $\delta = 3$ are depicted in Fig. 1.



(a) (b) (c)

Fig. 1. (a) Mask used for creating the synthetic IVUS data, (b) its corresponding IVUS B-mode reconstruction, and (c) recovered DBC values using 100 samplings ($\triangle P = 0.05$ *mm*, $\beta = 1$, and $\delta = 3$)

Real RF data: Experiments were performed using real IVUS RF data from six 40 *MHz* pullback sequences acquired with three different IVUS systems. These sequences correspond to different arteries from rabbits and swines. For each sequence we employed

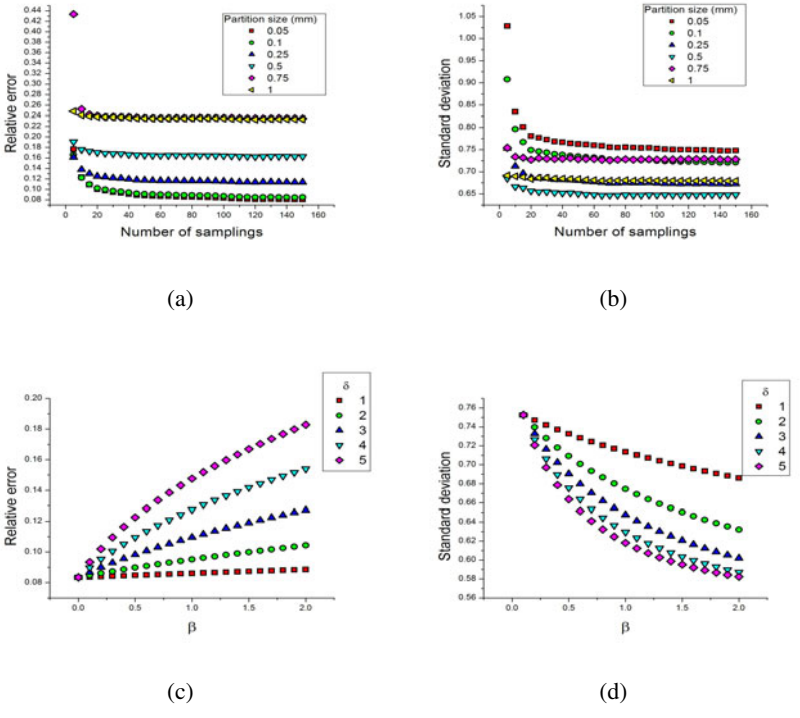


Fig. 2. (a, c) Relative error and (b, d) standard deviation for the mean of the recovered DBC values for two type of scatterers with respect to (a) the size of partition and (b) number of samplings, and (c) the cardinality of neighbors and (d) the value of the regularization term

our method on ten frames from different parts of the sequence and we compared the recovered DBC values for samples of lumen (i.e., blood) acquired from manual annotations provided by an expert. For these experiments the value of the width of the envelope of the impulse function was set to $\sigma = 5.3e^{-8}$, while the attenuation coefficient was set to the attenuation coefficient of blood (i.e., $\mu = 0.08276$ *dB/mm* [17]). The speed of sound was set to the speed of sound in a biological tissue ($c = 1540 \times 10^3$ *mm/s*). The size of partition was set to $\triangle P = 0.05$ *mm*, the density was set to $D = 400$ scatterers/mm^{-2} using the voxel approach of Rosales *et al.* [13]. The cardinality of the neighbors was set to $\delta = 3$ and $\beta = 1$. The information of the sequences and the mean of the recovered DBC values for each of the six cases are listed in Table 1.

As a preliminary blood detection experiment, we used our method to recover the DBC values from the IVUS RF data of a frame corresponding to a 40 *MHz* IVUS from swine (Fig. 3(a)), for which histological information is available. The regions corresponding to vascularization in the histology data have been annotated by an expert. The resulting DBC values for each pixel of the corresponding B-mode image are depicted using a color palette. Additionally, we created a pseudo-colored version of the IVUS image (Fig. 3(d)) using the DBC values and the same color representation. For

Table 1. Information of the sequences used for the experiments, and mean recovered DBC values for blood (τ_b) for each sequence

ID	Specimen	Artery	IVUS system	$\tau_b \times 10^{-3}$ (mm^2)	std $\times 10^{-3}$
1	Swine	Illiac	1	5.2	0.59
2	Swine	RCA	1	5.5	1.84
3	Swine	LAD	2	6.4	0.28
4	Swine	LAD	2	5.7	0.30
5	Rabbit	Aorta	3	5.5	0.56
6	Rabbit	Aorta	3	6.0	1.7

comparison, we have manually annotated the regions of the resulting image that correspond to vascularization based to the criterion that a vessel should contain a region with DBC values corresponding to blood surrounded by DBC values corresponding to non-blood. In Fig. 3(c) it can be observed that the recovered DBC values corresponding to blood from the lumen and the vessels in the adventitia are very similar. However, although there might be a correspondence between the VV in plaque (indicated by the blue rectangle in Fig. 3(b)) and some of the regions inside the plaque in the reconstructed images, we consider that the DBC might not be sufficient to detect such small vasculature by itself and should be considered along with other features.

4 Discussion

The size of the partition is a parameter that determines the size of the smallest structure that we can detect with our method. Moreover, from the synthetic data results it can be observed that, as the size of the partition decreases, the error between the recovered and true DBC values also decreases. However, the variability of the recovered DBC values increases as the size of the partition decreases. This variability is compensated with the regularization term. In the experiments with real data, the recovered DBC values for blood are similar for all the cases, which is an indication of the feasibility of using this approach for blood detection. By using a color map with the recovered DBC values, it is possible to generate pseudo colored IVUS images as the example depicted in Fig. 3(d), which may help the physicians to easier identify the different vessel structures. Exact correspondence between a histological slide and an IVUS image is difficult due to the variability in the orientation and position of the catheter and transducer. However, a fair correspondence may be achieved by locating large structures (side vessels) as in our histological example. While Figs. 3 (a, b, c) offer evidence of the feasibility of our method for extra-luminal blood detection, the manual detection of small vasculature such as VV in the recovered DBC or the colored B-mode images remains a difficult task. Therefore, a limitation of the present method is the lack of an automatic method for detecting the vasculature. We believe that the extra-luminal blood detection method can be improved by using the recovered DBC along with other image or RF-based features.

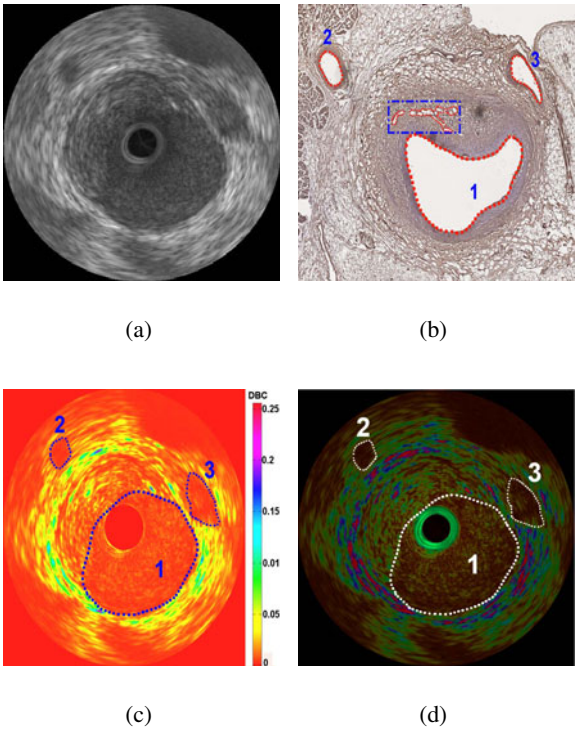


Fig. 3. (a) B-mode Cartesian frame of a 40 *MHz* IVUS swine case, (b) its corresponding histology, (c) recovered DBC values, and pseudo colored B-mode image obtained using the recovered DBC values. The main vasculature has been annotated by an expert observer on (b, c, and d).

5 Conclusions

We have presented a new method that employs a physics-based model of the IVUS RF signal for the computation of the DBC of the scatterers that generates the IVUS RF data. Our results are very encouraging and we believe that further research in this direction will lead to the development of a fast and reliable method for extra-luminal blood detection. Future work includes the use of overlapping partitions, additional quantitative validation, a method for the automatic segmentation of the vasculature, improvements to the scatterer model (i.e., adding attenuation by absorption), and the use of machine learning techniques for automatic blood detection.

References

1. Ciompi, F., Pujol, O., Leor, O., Gatta, C., Vida, A., Radeva, P.: Enhancing in-vitro IVUS data for tissue characterization. In: Proc. 4th Iberian Conference on Pattern Recognition and Image Analysis, Povoa de Varzim, Portugal, June 10-12, pp. 241–248 (2009)
2. Fontaine, I., Bertrand, M., Cloutier, G.: A system-based approach to modeling the ultrasound signal backscattered by red blood cells. Biophysical Journal 77(5), 2387–2399 (1999)

3. Heistad, D.D., Marcus, M.L.: Role of vasa vasorum in nourishment of the aorta. Blood Vessels 16, 225–238 (1979)
4. Katouzian, A., Baseri, B., Konofagou, E.E., Laine, A.F.: An alternative approach to spectrum-based atherosclerotic plaque characterization techniques using intravascular ultrasound (IVUS) backscattered signals. In: Proc. 2nd MICCAI Workshop on Computer Vision for Intravascular and Intracardiac Imaging, New York, NY (2008)
5. Kawasaki, M., Takatsu, H., Noda, T., Sano, K., Ito, Y., Hayakawa, K., Tsuchiya, K., Arai, M., Nishigaki, K., Takemura, G., Minatoguchi, S., Fujiwara, T., Fujiwara, H.: In vivo quantitative tissue characterization of human coronary arterial plaques by use of integrated backscatter intravascular ultrasound and comparison with angioscopic findings. Circulation 105, 2487–2492 (2002)
6. Koga, T., Uchino, E., Tanaka, Y., Suetake, N., Hiro, T., Matsuzaki, M.: Tissue characterization of coronary plaque by using fractal analysis-based features of IVUS RF-signal. In: Proc. 5th International Workshop on Computational Intelligence and Applications, Hiroshima, Japan, November 10-12, pp. 77–81 (2009)
7. Langheinrich, A.C., Michniewicz, A., Sedding, D.G., Walker, G., Beighley, P.E., Rau, W.S., Bohle, R.M., Ritman, E.L.: Correlation of vasa vasorum neovascularization and plaque progression in aortas of apolipoprotein E(-/-)/low-density lipoprotein(-/-) double knockout mice. Arteriosclerosis, Thrombosis, and Vascular Biology 26(2), 347–352 (2006)
8. Maton, A.: Human biology and health. Prentice Hall, Englewood Cliffs (1993)
9. Mendizabal-Ruiz, E., Biros, G., Kakadiaris, I.A.: An inverse scattering algorithm for the segmentation of the luminal border on intravascular ultrasound data. In: Proc. 12th International Conference on Medical Image Computing and Computer Assisted Intervention, London, UK, September 20- 24, pp. 885–892 (2009)
10. Nair, A., Kuban, B., Tuzcu, E., Schoenhagen, P., Nissen, S., Vince, D.: Coronary plaque classification with intravascular ultrasound radiofrequency data analysis. Circulation 106(17), 2200–2206 (2002)
11. 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)
12. O'Malley, S.M., Naghavi, M., Kakadiaris, I.A.: One-class acoustic characterization applied to blood detection in IVUS. In: Proc. 10th International Conference on Medical Image Computing and Computer Assisted Intervention, Brisbane, Australia, October 29 - November 2, pp. 202–209 (2007)
13. Ramirez, M., Radeva, P., Mauri, J., Pujol, O.: Simulation model of intravascular ultrasound images. In: Proc. 7th International Conference on Medical Image Computing and Computer-Assisted Intervention, Saint-Malo, France, September 26 - 30, pp. 200–207 (2004)
14. Reeder, B.: Acoustic scattering by axisymmetric finite-length bodies with application to fish: measurement and modeling. Ph.D. thesis, Massachusetts Institute of Technology (2002)
15. Roodaki, A., Taki, A., Setarehdan, S.K., Navab, N.: Modified wavelet transform features for characterizing different plaque types in IVUS images; a feasibility study. In: Proc. 9th International Conference on Signal Processing, Beijing, China, October 26 - 29, pp. 789–792 (2008)
16. Ross, R.: The pathogenesis of atherosclerosis: A perspective for the 1990s. Nature 362, 801–809 (1993)
17. Shung, K.K., Smith, M.B., Tsui, B.: Principles of medical imaging. Academic Press, London (1992)
18. Thijssen, J., Oosterveld, B.: Perfomance of echographic equipment and potentials for tissue characterization. Mathematics and Computer Science in Medical Imaging F39, 455–468 (1998)

Physiologically Based Construction of Optimized 3-D Arterial Tree Models*

Matthias Schneider[1], Sven Hirsch[1], Bruno Weber[2], and Gábor Székely[1]

[1] Computer Vision Laboratory, ETH Zurich, Switzerland
[2] Institute of Pharmacology and Toxicology, University of Zurich, Zurich, Switzerland

Abstract. We present an approach to generate 3-D arterial tree models based on physiological principles while at the same time certain morphological properties are enforced at construction time in order to build individual vascular models down to the capillary level. The driving force of our approach is an angiogenesis model incorporating case-specific information about the metabolic activity in the considered domain. Additionally, we enforce morphometrically confirmed bifurcation statistics of vascular networks. The proposed method is able to generate artificial, yet physiologically plausible, arterial tree models that match the metabolic demand of the embedding tissue and fulfill the enforced morphological properties at the same time. We demonstrate the plausibility of our method on synthetic data for different metabolic configurations and analyze physiological and morphological properties of the generated tree models.

Keywords: vascular tree construction, arterial tree model, angiogenesis, morphological statistics, computer simulation, computational physiology.

1 Introduction

An in-depth understanding of the microvascular structure is required in many research areas both in normal and pathological tissue. As an example, detailed, explicit vascular models are needed in surgical training simulation, when generating variable anatomical scenes with realistic physiological properties [14]. Similarly, knowledge of the entire cerebrovascular network down to capillary level is required for gaining insight into blood flow dynamics and its regulation by numerical simulations [10]. Additionally, it has been shown that many of the neurodegenerative diseases (*e.g.* Alzheimer's disease) have a prominent vascular component, and there is increasing evidence that reduced energy substrate and oxygen delivery is in part responsible for the severe symptoms of the disease. This underlines the need for a better knowledge of the vascular network's structure in normal and pathological tissue in order to increase our understanding of the pathophysiological mechanisms. Microvascular structures have been analyzed based on intravascular dye injections, staining of vessel components, or vascular corrosion casts using state of the art imaging modalities, *e.g.*, all-optical histology, or synchrotron radiation based x-ray tomographic microscopy. Since the segmentation and reconstruction of consistent arterial trees still remains a challenge, numerous methods for the generation of artificial vascular models have been proposed.

* Supplementary material for this article is available at http://www.vision.ee.ethz.ch/ReCoVa

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 404–411, 2011.

The main approaches are typically based on optimality or evolutionary principles. Nekka *et al.* [9] proposed a deterministic construction method for 2-D vascular structures incorporating a highly simplified angiogenesis model. Vascular formation is modeled as evolutionary process in response to angiogenic factors produced by ischemic tissue cells. As the formation process is purely based on the distribution of angiogenic factors without imposing any further restriction regarding the geometry of the evolving network, the results appear rather artificial and too regular. A more sophisticated angiogenesis-based simulation framework for the construction of vascular systems in arbitrary anatomies has been proposed by Szczerba *et al.* [13]. Even though the results show high similarity with real vasculatures, the underlying simplifications still do not allow to generate vascular morphologies in full accordance with experimental findings.

Based on experimental observations from real vessel networks, several optimality principles have long been hypothesized, *e.g.*, minimal building material or minimal energy dissipation [8,15]. Klarbring *et al.* [6] apply strategies for topology optimization of electrical networks and load carrying trusses to fluid mechanics. The flow network topology is optimized w.r.t. the cross-section of the pipes in order to minimize the total pressure loss (dissipation) under a total volume constraint.

Schreiner and Buxbaum [11] proposed to use constrained constructive optimization (CCO) to iteratively construct a 2-D binary tree. The method of CCO (and its variations) is a stochastic construction process purely based on structural optimality principles under hemodynamic boundary conditions. In order to generate visually more realistic vascular trees in 3-D, Karch *et al.* [5] combined CCO with staged tissue growth leading to structural changes of the simulated tree models. However, the vascular networks have been truncated at the pre-arteriolar scale. Similarly, CCO has recently been combined with time-dependent constraints on a level set distance function to restrict the morphology of major (cerebral) arteries, in particular [1].

In contrast, we pursue a modeling approach based on physiological principles while enforcing certain morphological properties at construction time in order to build individual vascular models down to the capillary level. In our method, we use an *angiogenesis model* as the driving force of vascular tree formation while relying on the metabolic activity of the tissue in the considered domain as may be obtained from functional imaging, for instance. At the same time we incorporate morphometrically confirmed optimality hypotheses concerning the branching structure of vascular systems. This way, we iteratively construct an arterial tree model that meets the metabolic requirements of the embedding tissue on a gradually growing domain. As opposed to previously suggested approaches, we do not explicitly enforce hemodynamic constraints.

2 Methods

Our approach for the generation of artificial arterial trees is based on physiological principles related to (sprouting) angiogenesis in order to achieve physiologically plausible results. Angiogenesis describes the formation of new capillary blood vessels from a pre-existing vasculature in well-characterized stages. It plays a crucial role in different growth processes such as embryonic development, wound healing, or tumor growth [7]. In chemotactic response to angiogenic signals, also known as angiogenic growth factors, solid capillary sprouts develop from pre-existing vessels by means of endothelial

cell (EC) proliferation and migration towards the source of the angiogenic stimulus. We will reduce the complex signaling cascade involved in angiogenesis to a single transmitter, namely the vascular endothelial growth factor (VEGF), which has widely been studied and proven to be a potent stimulator of physiological and pathological angiogenesis [3]. VEGF, secreted by, *e.g.*, tumor or ischemic cells, diffuses into the surrounding tissue establishing a concentration gradient between the source and the vessels. In our method, these physiological principles are applied on a gradually growing simulation domain to drive the iterative construction of an arterial tree model.

Arterial Tree Model. In this work, we consider arterial trees rather than full-fledged vascular networks. According to morphometric analysis, bifurcations of vascular trees almost invariably branch into two distal branches [15]. Arterial trees can hence be considered as binary trees. Each vessel segment is represented by an edge which is modeled as rigid cylindrical tube with radius r and length l connecting two nodes. This gives rise to four different types of nodes, namely root, leaf, bifurcation (⊸), and inter nodes (—). The latter links two successive segments and allows to approximate the shape of a tortuous vessel branch. Note that this model does not allow for interconnections (anastomoses). In the following, we will distinguish between tree *segments* (single edge between any two connected nodes) and *branches* (path from a bifurcation or root node to the next distal bifurcation or leaf node).

The relation of the decreasing vessel radii from proximal to distal segments is governed by a *bifurcation law* (Murray's Law) relating the radius of the proximal parent branch r_p to the radii r_l, r_r of the left and right daughter branches [8]: $r_p^\gamma = r_l^\gamma + r_r^\gamma$, where γ denotes the bifurcation exponent, with values reported ranging from $\gamma = 2.0$ to $\gamma = 3.0$ [2,11].

The bifurcation geometry is further constrained w.r.t. the *bifurcation angles* based on fluid dynamic considerations [2]:

$$cos(\phi_l) = \frac{r_p^4 + r_l^4 - r_r^4}{2r_p^2 r_l^2}\,, \quad cos(\phi_r) = \frac{r_p^4 + r_r^4 - r_l^4}{2r_p^2 r_r^2} \quad , \tag{1}$$

with ϕ_l, ϕ_r denoting the bifurcation angle of the left and right daughter branch, respectively. Geometrically, this corresponds to the optimal position of the branching point $\boldsymbol{p}_b$ minimizing the total lumen volume for fixed boundary points:

$$\hat{\boldsymbol{p}}_b = \underset{\boldsymbol{p}_b}{\operatorname{argmin}} \sum_{k\in\{p,l,r\}} r_k^2 \left\|\boldsymbol{p}_k - \boldsymbol{p}_b\right\|_2 \quad , \tag{2}$$

where $\boldsymbol{p}_p$, $\boldsymbol{p}_l$, $\boldsymbol{p}_r$ are the boundary nodes of the parent, left, and right segment, respectively (see Figure 1a).

Angiogenesis Model. We use a simplified angiogenesis model as the driving force for our simulations. The model considers the mutual interplay of arterial oxygen (O_2) supply and VEGF secreted by ischemic cells. Tissue is assumed to be homogeneous w.r.t. O_2 and VEGF transport with diffusivity D_1 and D_2, respectively. Assuming steady-state conditions, Fick's first law postulates [12]:

$$D_1 \nabla^2 c_1 = R_1(c_1)\,, \quad R_1(c_1) = R_1^0 \frac{c_1}{c_1 + c_1^h} \quad , \tag{3}$$

where c_1 represents the O_2 concentration. $R_1(c_1)$ denotes the O_2 consumption rate which is described by a Michaelis-Menten relationship defined by the O_2 demand at saturation level R_1^0 (unlimited O_2 supply) and the half-saturation concentration c_1^h. Note that R_1^0 is subject to regional variations, in general.

VEGF secretion in tissue is dependent on the average oxygenation level. However, the exact relationship has not yet been experimentally measured. We model the oxygen-dependent VEGF secretion rate as follows [3]:

$$R_2(c_1) = \begin{cases} 6R_2^0 & , c_1 \leq t_1^{\text{low}} \\ (1+5\left(\frac{t_1^{\text{high}}-c_1}{t_1^{\text{high}}-t_1^{\text{low}}}\right)R_2^0 & , t_1^{\text{low}} < c_1 < t_1^{\text{high}} \\ R_2^0 & , c_1 \geq t_1^{\text{high}} \end{cases} . \tag{4}$$

The secreted VEGF isotropically diffuses and is subject to natural decay [7]:

$$D_2\nabla^2 c_2 = \nu_2 c_2 - R_2(c_1) \quad , \tag{5}$$

with diffusivity D_2 and decay rate ν_2, while neglecting VEGF uptake by EC.

Vascular Growth. Vascular growth is modeled as a chemotactic process w.r.t. the VEGF concentration. In general, we distinguish between *apical growth* at leaf nodes (apices) and *sprouting* at inter nodes. A capillary may either elongate or bifurcate into two similar branches resulting mostly in symmetric bifurcations. This decision is made randomly with a sigmoidal bifurcation probability $p_b(a) = \text{sig}\,(-\eta(1-a/A_0))$, where $a = l/r$ denotes the aspect ratio of the proximal branch with length l and radius r. The probability distribution $p_b(a)$ can be considered as global bifurcation length statistics with adjustable shape and normalization factor η and A_0, respectively. In case of elongation at a leaf node, the radius of the new distal segment is adopted from the proximal segment. The growth direction $\boldsymbol{d}_g$ is estimated by $\boldsymbol{d}_g = \nabla c_2(\boldsymbol{x})/\|\nabla c_2(\boldsymbol{x})\|_2 + \lambda_g \boldsymbol{d}_s$, where $\boldsymbol{d}_s$ denotes the (normalized) direction of the proximal segment and λ_g is a constant weighting factor. The distal segment length and the lengths and radii of the two new branches are drawn from normal distributions, while enforcing a minimum radius constraint $r \geq r_{\min}$. According to the principle of minimal building material, the parent and daughter segments at a bifurcation node are coplanar [2]. The branching angles are computed according to Equation (1). The two distal daughter branches are arranged such that the angle between the VEGF gradient and the growth direction of either of the two segments is minimized (see Figure 1a).

Besides apical growth at leaf nodes, vessels may also build sprouts at inter nodes producing both symmetric and asymmetric bifurcations. The sprouting probability is defined as $p_s = \min\left(p_b(l_p/r), p_b(l_d/r)\right)$, where r, l_p, and l_d denote the radius and length of the proximal and distal branch meeting at the inter node. Similar to the bifurcation case, the length and radius of the new sprout are normally distributed. The VEGF gradient determines the sprout direction.

Both sprouting inter nodes and bifurcating leaf nodes spoil the bifurcation law, in general. Additional tree rebalancing is required to reinforce optimality. To this end, the segment radii have to be updated along the path from the site of growth up to the root node according to Murray's Law. Adjusting the radii, in turn, impairs the optimality of the branching angles. However, the bifurcations can be rebalanced by relocating the bifurcation node according to Equation (2).

Table 1. Main simulation parameters

Param	Value	Unit	Ref
D_1	$2.41 \cdot 10^{-9}$	$\mathrm{m^2 s^{-1}}$	[4]
R_1^0	$3.01 \cdot 10^{-3}$	$\mathrm{m^3_{O_2} m^{-3}_{tissue} s^{-1}}$	[4]
c_1^h	$1.95 \cdot 10^{-5}$	$\mathrm{m^3_{O_2} m^{-3}_{tissue}}$	[4]
D_2	$1.04 \cdot 10^{-10}$	$\mathrm{m^2 s^{-1}}$	[3]
ν_2	$1.81 \cdot 10^{-4}$	$\mathrm{s^{-1}}$	[7]
R_2^0	$2.50 \cdot 10^{-18}$	$\mathrm{mol\, m^{-2} s^{-1}}$	[3]
t_1^{low}	$3.89 \cdot 10^{-5}$	$\mathrm{m^3_{O_2} m^{-3}_{tissue}}$	[3]
t_1^{high}	$7.78 \cdot 10^{-4}$	$\mathrm{m^3_{O_2} m^{-3}_{tissue}}$	[3]
γ	3	-	[2]
θ_1	$1.5 \cdot R_1^0$	$\mathrm{m^3_{O_2} m^{-3}_{tissue} s^{-1}}$	-
η	7	-	-
A_0	10	-	-
λ_g	1	-	-
$r_{\min}$	$3.50 \cdot 10^{-6}$	m	-

Algorithm 1. Arterial tree construction

Require: domain Ω_0, tree(s) Ψ_0, metabolic map $\mathcal{R}_0$
  for scale $s = 1$ **to** N **do**
    $[\Omega_s, \mathcal{R}_s, \Psi_s] \leftarrow$ rescale $\Omega_{s-1}, \mathcal{R}_{s-1}, \Psi_{s-1}$
    repeat
      $c_1 \leftarrow O_2$-map(Ψ_s,Ω_s,R_s)
      $c_2 \leftarrow$ VEGF-map(c_1,Ω_s)
      $\Psi_s \leftarrow$ grow-tree(Ψ_s,c_2)
      $\Psi_s \leftarrow$ rebalance(Ψ_s)
    until supply rate $\geq \kappa_s$
  end for

Table 2. Global quantities (mean±std)

total	$\mathcal{M}_1$	$\mathcal{M}_2$
length [m]	1.94 ± 0.018	2.68 ± 0.009
surface [$\mathrm{mm^2}$]	166 ± 1.00	214 ± 1.06
volume [$\mathrm{mm^3}$]	1.89 ± 0.070	2.25 ± 0.037
# terminals	$1.25 \cdot 10^4 \pm 132$	$1.84 \cdot 10^4 \pm 119$

Iterative Tree Construction. We use a multiscale approach for the iterative tree construction as summarized in Algorithm 1. The scaling can be considered as virtual staged growth of the domain (tissue) from a downscaled domain Ω_1 ($s = 1$) to the final target domain $\Omega_N = \Omega_0$ ($s = N$). For a given arterial tree and scale s, the total oxygenation of the simulation domain Ω_s is computed as the superimposition of the oxygenation maps of each individual tree segment according to Equation (3) assuming radial O_2 diffusion with a Dirichlet boundary condition on the surface of the artery wall. The O_2 concentration inside the vessel lumen is in fact considered constant $c_1^l = 2.04 \cdot 10^{-4}\, \mathrm{m^3_{O_2} m^{-3}_{blood}}$, assuming 100% O_2 saturation of hemoglobin at a concentration of $1.50 \cdot 10^4\, \mathrm{g_{Hb}\ m^{-3}_{blood}}$, and 95 mmHg partial pressure of O_2 in blood. The O_2 concentration on the surface of the artery wall has been chosen as $25\%\, c_1^l$ which corresponds to a maximum O_2 diffusion distance in tissue of approximately $100\, \mu\mathrm{m}$ [12]. The induced VEGF steady-state concentration map is subsequently computed according to Equation (5) and used as an excitatory potential field. The tree nodes are sampled "in rounds" with random order within each round to find excited nodes. A node is considered excited if the local VEGF level exceeds a threshold θ_1. After the tree has grown at an excited node, the O_2 and VEGF maps are updated accordingly. The simulation proceeds to the next scale if the O_2 supply rate exceeds a scale-dependent level κ_s.

3 Simulation Results

We have tested and analyzed the iterative tree construction approach described in Algorithm 1 for different synthetic data sets. A cylindrical simulation domain Ω_0 of height and diameter 4.1 mm has been discretized by a grid with isotropic spacing $32\, \mu\mathrm{m}$ (128^3 voxels). The initial trees Ψ_0 have been defined by six tiny sprouts ($r = 5\, \mu\mathrm{m}$) on the surface of the cylinder each marking an entry point of a feeding vessel. We have investigated two different configurations for the prescribed metabolic map $\mathcal{R}_0$. Model $\mathcal{M}_1$ assumes uniform metabolic activity $\mathcal{R}_0 = R_1^0$. The metabolic map of model $\mathcal{M}_2$ shows an increased level of O_2 consumption at the center of the domain. The maximum level

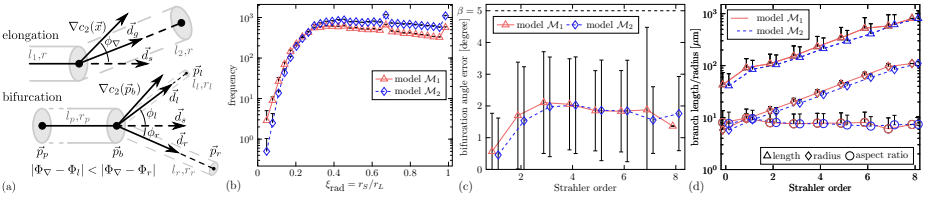


Fig. 1. (a) Apical growth at leaf node. (b) Average frequency of bifurcation symmetry index ξ_{rad}. (c) Average bifurcation angle deviation over diameter-defined Strahler order of the parent segment [5]. The correction threshold β is indicated by the dashed line. (d) Average length, radius, and aspect ratio of branches over Strahler order. Upward error bars indicate the standard deviation over all eight simulations per model.

at the center has been set to $11 \cdot R_1^0$ and decays to R_1^0 according to a Gaussian profile as shown in Figure 3a. The simulations have been conducted for $N = 32$ scales starting at a scaling factor of $1/16$ corresponding to an effective voxel spacing of $2\ \mu\text{m}$. The supply rate has been computed as the ratio of perfused voxels and the total number of voxels in the domain. A voxel was considered perfused if the effective O_2 concentration exceeded t_1^{high}, *i.e.*, there was no extra VEGF secreted. The scale-dependent target supply rate κ_s was linearly interpolated between 50% ($s = 1$) and 95% ($s = N$). Defective bifurcation configurations have been readjusted by tree remodeling if the maximum deviation of the left and right bifurcation angle from the optimal bifurcation angle exceeded $\beta = 5°$.

We have simulated eight realizations of each model $\mathcal{M}_1$ and $\mathcal{M}_2$ using different seeds for the employed pseudo random number generator (PRNG). The simulation parameters are summarized in Table 1. Figure 2 visualizes the evolving vasculature at different scales for a single realization of $\mathcal{M}_2$ as well as the final tree model of $\mathcal{M}_1$.

Despite remarkable variations of the visual appearance of the constructed tree models depending on the PRNG seeds, global geometric quantities such as total lumen volume, surface, and total segment length remained surprisingly stable with virtually no variation for different seeds as summarized in Table 2. However, there are significant differences between the simulation models. The trees constructed for $\mathcal{M}_2$ show increased figures w.r.t. the considered global geometric properties. This can be explained by the fact that the trees have to deliver much more oxygen to the tissue corresponding to the increased level of metabolic activity.

We have analyzed the bifurcation pattern in terms of the local symmetry index $\xi_{\text{rad}} = r_S/r_L$, where r_S and r_L denote the radius of the smaller and larger distal segment, respectively [5]. The estimated distribution of ξ_{rad} is shown in Figure 1b. Comparing the bifurcation symmetry, we note that the trees of $\mathcal{M}_1$ show an increased number of asymmetric bifurcations (conveying vessels for transport of blood across larger distances), whereas the vasculature of $\mathcal{M}_2$ is clearly dominated by symmetric bifurcations (mostly resulting from delivering vessels). The "harmonic" peaks at $\xi_{\text{rad}} = 4/9, 2/3, 1$ result from a constraint on the choice of bifurcation radii (radii clamped to $[r_{\min}, 3/2 \cdot r_p]$) to avoid excessive enlargement of the proximal segment.

Tree remodeling, which had become necessary to rebalance the constructed tree after sprouting or bifurcating, has been validated w.r.t. the optimality of bifurcation angles. To this end, the average deviation of the observed bifurcation angles from the optimal

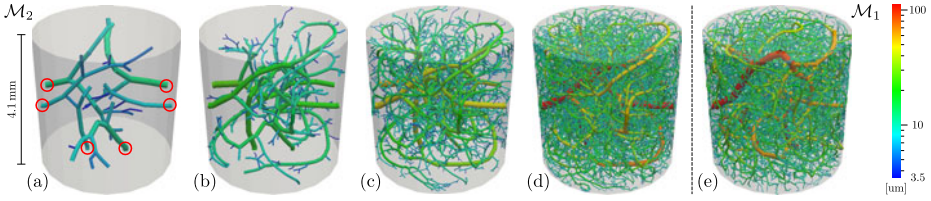


Fig. 2. Evolution of constructed trees for model $\mathcal{M}_2$ at scales $s = 8, 16, 24, 32$ (a-d). (e) Arterial tree model for model $\mathcal{M}_1$ at scale $s = 32$. The downscaled simulation domains have been normalized to the size of the target domain (d,e) for better comparability. The segment radii are color-coded on a logarithmic scale. The entry points of the feeding vessels are marked red (a).

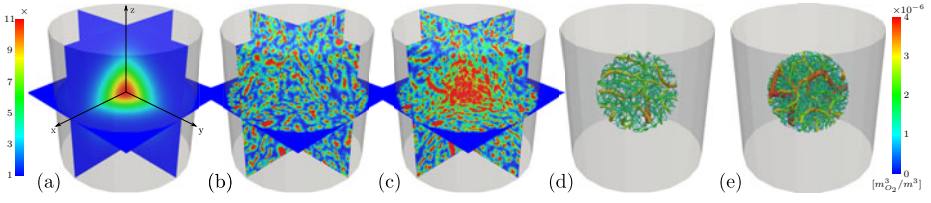


Fig. 3. (a) Synthetic metabolic activity used for model $\mathcal{M}_2$. (b,c) Simulated perfusion maps for arterial trees of model $\mathcal{M}_1$ (b) and $\mathcal{M}_2$ (c). (d,e) Arterial tree models within a centered sphere of radius 1.2 mm for $\mathcal{M}_1$ (d) and $\mathcal{M}_2$ (e). The segment radii are color-coded as in Figure 2.

configuration has been computed for each bifurcation as shown in Figure 1c. The average angular variations stay well below the correction threshold $\beta = 5\,^\circ$ that has been used in the simulations.

It is also interesting to analyze the trees w.r.t. the bifurcation length statistics that has been used for the construction to enforce a certain morphological structure. To this end, we have classified the branches by Strahler order and computed the average length and radius for all classes as reported in Figure 1d. Length and radius show an exponential profile for both models. The average aspect ratio is almost constant over all Strahler orders at a level of about 10 which corresponds to the normalization factor A_0 that has been used for our simulations and hence matches the "expected" value ($p_b(A_0) = 0.5$).

The perfusion maps of the artificial tree models are visualized in Figure 3. As opposed to $\mathcal{M}_1$ with uniform metabolic activity, the perfusion map for $\mathcal{M}_2$ increases towards the center of the domain to meet the prescribed non-uniform metabolic demand. As already indicated by the bifurcation symmetry analysis, the vascular density in the center of the domain is significantly higher for $\mathcal{M}_2$ as opposed to $\mathcal{M}_1$ (see Figure 3d-e).

Explicit comparison of the tree models with real vascular systems is still an open problem. Considering the structural complexity and the inherently stochastic nature of vessel formation, particularly at the capillary level, comparing functional properties and structural statistics might be the best approach to define meaningful similarity measures.

4 Conclusions and Future Work

We presented an iterative multi-scale approach for the construction of optimized 3-D arterial tree models based on physiological principles. The driving force of our

simulations is an angiogenesis model using individual, case-specific information about the metabolic activity of the tissue in the considered domain. Besides, bifurcation statistics and morphological optimality principles are incorporated into the construction. We presented promising results for different synthetic datasets, demonstrating the ability of the algorithm to respect prescribed morphological constraints.

As for future work, there are still many unconsidered aspects, *e.g.*, transition from vascular trees to network-like structures including anastomoses. Likewise, it would be interesting to analyze the fluid dynamic properties of the generated tree models in more detail. The most significant challenge we face is the comparison of the results with experimentally observed real vascular systems w.r.t. structural and functional properties.

Acknowledgements. This work has been funded by the Swiss National Centre of Competence in Research on Computer Aided and Image Guided Medical Interventions (NCCR Co-Me) funded by the Swiss National Science Foundation.

References

1. Bui, A.V., Manasseh, R., Liffman, K., Sutalo, I.D.: Development of optimized vascular fractal tree models using level set distance function. Med. Eng. Phys. 32(7), 790–794 (2010)
2. Fung, Y.C.: Biomechanics, 2nd edn. Springer, New York (1997)
3. Gabhann, F.M., Ji, J.W., Popel, A.S.: VEGF gradients, receptor activation, and sprout guidance in resting and exercising skeletal muscle. J. Appl. Physiol. 102(2), 722–734 (2007)
4. Ji, J.W., et al.: A computational model of oxygen transport in skeletal muscle for sprouting and splitting modes of angiogenesis. J. Theor. Biol. 241(1), 94–108 (2006)
5. Karch, R., Neumann, F., Neumann, M., Schreiner, W.: Staged growth of optimized arterial model trees. Ann. Biomed. Eng. 28(5), 495–511 (2000)
6. Klarbring, A., Petersson, J., Torstenfelt, B., Karlsson, M.: Topology optimization of flow networks. Comput. Method Appl. M 192(35-36), 3909–3932 (2003)
7. Milde, F., Bergdorf, M., Koumoutsakos, P.: A hybrid model for three-dimensional simulations of sprouting angiogenesis. Biophys. J. 95(7), 3146–3160 (2008)
8. Murray, C.D.: The physiological principle of minimum work: I. The vascular system and the cost of blood volume. Proc. Natl. Acad. Sci. USA 12(3), 207–214 (1926)
9. Nekka, F., Kyriacos, S., Kerrigan, C., Cartilier, L.: A model of growing vascular structures. Bull. Math. Biol. 58(3), 409–424 (1996)
10. Reichold, J., et al.: Vascular graph model to simulate the cerebral blood flow in realistic vascular networks. J. Cereb. Blood Flow Metab. 29(8), 1429–1443 (2009)
11. Schreiner, W., Buxbaum, P.F.: Computer-optimization of vascular trees. IEEE Trans. Biomed. Eng. 40(5), 482–491 (1993)
12. Secomb, T.W., et al.: Green's function methods for analysis of oxygen delivery to tissue by microvascular networks. Ann. Biomed. Eng. 32(11), 1519–1529 (2004)
13. Szczerba, D., Székely, G.: Simulating vascular systems in arbitrary anatomies. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3750, pp. 641–648. Springer, Heidelberg (2005)
14. Tuchschmid, S., Bajka, M., Szczerba, D., Lloyd, B.A., Székely, G., Harders, M.: Modelling intravasation of liquid distension media in surgical simulators. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part I. LNCS, vol. 4791, pp. 717–724. Springer, Heidelberg (2007)
15. Zamir, M.: Optimality principles in arterial branching. J. Theor. Biol. 62(1), 227–251 (1976)

Cortical Surface Strain Estimation Using Stereovision

Songbai Ji[1], Xiaoyao Fan[1], David W. Roberts[2,3], and Keith D. Paulsen[1,2]

[1] Thayer School of Engineering, Dartmouth College, Hanover, NH 03755
[2] Norris Cotton Cancer Center, Lebanon, NH 03756
[3] Dartmouth Hitchcock Medical Center, Lebanon, NH 03756
{songbai.ji,xiaoyao.fan,david.w.roberts,
keith.d.paulsen}@dartmouth.edu

Abstract. We present a completely noninvasive technique to estimate soft tissue surface strain by differentiating three-dimensional displacements obtained from optical flow motion tracking using stereo images. The implementation of the strain estimation algorithm was verified with simulated data and its application was illustrated in three open cranial neurosurgical cases, where cortical surface strain induced by arterial blood pressure pulsation was evaluated. Local least squares smoothing was applied to the displacement field prior to strain estimation to reduce the effect of noise during differentiation. Maximum principal strains (ε^1) of up to 7% were found in the exposed cortical area on average, and the largest strains (up to ~18%) occurred near the craniotomy rim with the majority of ε^1 perpendicular to the boundary, indicating relative stretching along this direction. The technique offers a new approach for soft tissue strain estimation for the purpose of biomechanical characterization.

1 Introduction

Tracking cortical surface motion noninvasively using stereovision is a practical and important intraoperative imaging technique to compensate for brain shift in image-guided neurosurgery. Typically, cortical surface displacement data is obtained through a rigid or nonrigid registration between the reconstructed stereo surface and preoperative magnetic resonance images [1] or between two stereo surfaces acquired at different surgical stages [2]. Because the parenchyma is constantly subjected to complex loading and boundary conditions throughout surgery (e.g., from surgical intervention, confinement by the craniotomy, and changes in intracranial pressure), the brain often deforms nonrigidly within the field of view. Although considerable attention has been focused on tracking cortical surface displacements, little effort has been devoted to investigating the non-uniformity of their *distribution* that directly determines the tissue surface strain.

Tissue surface strain is an important indictor of inherent tissue compliance and has been investigated in studies of the biomechanics of the eye where principal strain of the scleral shell detected from a single digital camera was found to correlate inversely with intraocular pressure [3]. Because a more compliant material undergoes a larger strain under the same applied force, tissue surface strains directly reflect the level of

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 412–419, 2011.

tissue compliance, which is an important indicator of physiological state, especially in the brain (e.g., [4]). A comprehensive understanding of the state of tissue strain may also be clinically valuable when the risk of mechanical damage exists (e.g., in cases of herniation, tissue retraction, and decompressive craniectomy).

In this paper, we present a computational framework to estimate surface strain from tracked tissue motion using stereovision and illustrate its application in the context of open cranial neurosurgery. Specifically, we estimate dynamic cortical surface strains induced by arterial blood pressure pulsation after dural opening, which is commonly observed but has not been previously quantified. We focus on the essential computational aspects of the approach in order to establish a foundation for the technique in anticipation of future studies relating cortical surface strain to clinical factors such as brain compliance and/or risk of tissue damage.

2 Material and Methods

2.1 Cortical Surface Reconstruction through Stereovision

Acquisition of stereo image pairs was externally triggered using two charge-coupled device (CCD) cameras (image size of 768 × 1024; pixel resolution of approximately 50–100 μm; images are in red, green, and blue (RGB)) rigidly attached to a surgical microscope. The stereo image acquisition frame rate varied for different patients (4–10 frames per second (fps)) but remained relatively constant for a given subject. Techniques for stereo image calibration and reconstruction based on a pinhole camera model and radial lens distortion correction can be found, e.g., in [1], and are briefly outlined here for completeness. A 3D point in world space (X, Y, Z) is transformed into the camera image coordinates (x, y) using a perspective projection matrix:

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} \alpha_x & 0 & C_x & 0 \\ 0 & \alpha_y & C_y & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \times \mathbf{T} \times \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}, \tag{1}$$

where α_x and α_y incorporate the perspective projection from camera to sensor coordinates and the transformation from sensor to image coordinates, (C_x, C_y) is the image center, while $\mathbf{T}$ is a rigid body transformation describing the geometrical relationship between the two cameras. A total of 11 camera parameters (6 extrinsic: 3 rotation and 3 translation; and 5 intrinsic: focal length, f, lens distortion parameter, k_1, scale factor, S_x, and image center, (C_x, C_y)) are determined through calibration using a least squares approach.

Stereo matching was facilitated by constraining the search for correspondence points along an epipolar line (defined as the projection of the optical ray of one camera via the center of the other camera following a pinhole model) through image rectification and subsequent intensity correlation and smoothness criteria defined in [1]. Each pair of correspondence points was transformed into their respective 3D camera space using the intrinsic parameters, and then transformed into a common 3D space using the extrinsic parameters. Together with their respective camera centers in

the common space, two optical rays were constructed with their intersection defining the 3D location of the correspondence point pair.

2.2 Motion Tracking through Optical Flow

Because radial image distortion is effectively compensated for during rectification, rectified images of the left camera taken at times t and (t+1) were used to track surface displacement. Similarly to [3], displacements were tracked from continuously acquired RGB images using an optical flow algorithm subject to constraint:

$$I_x u_x + I_y u_y + I_t = 0 \ , \tag{2}$$

where I_x, I_y and I_t are the derivatives of RGB image intensity at spatio-temporal location (x, y, t) in the corresponding directions, and (u_x, u_y) denotes displacement in the x and y directions. We employed an algorithm proposed in [5] based on [6] in which a gradient constancy assumption (Eqn. 3) is applied to allow small variations in image gray values to determine displacement vectors:

$$\nabla I(x, y, t) = \nabla I\left(x + u_x, y + u_y, t + 1\right) . \tag{3}$$

In addition, a discontinuity-preserving spatio-temporal smoothness constraint was further applied to generate a piecewise smooth flow field. A multiscale approach starting with a coarse, smoothed image set is also used to ensure global minimization [6]. The computational cost to achieve full-field displacements for each pair of rectified images was approximately 70 sec. The computational efficiency can be further improved by down-sampling the input images (e.g., by selecting one of every three pixels in both directions for both images, leading to a computational cost of approximately 12 sec), which did not result in a significant change in displacement (<~5%). The start and end points of each displacement vector in image space were then transformed into 3D space through stereovision, and the corresponding displacement vector was generated by subtracting the corresponding point locations.

2.3 In-Plane Strain Estimation

In-plane strain was calculated numerically by differentiating the displacement field. Because of noise in the measured displacements, smoothing prior to differentiation is typical [7], and a point-wise local least squares fitting scheme was employed for this purpose. For each 3D point of interest, a linear plane was used to approximate displacements (u_x, u_y, u_z) from the neighboring set of points of size (2m+1)×(2m+1):

$$\begin{aligned} u_x(i, j) &= a_0 + a_1 x_{i,j} + a_2 y_{i,j} + a_3 z_{i,j} , \\ u_y(i, j) &= b_0 + b_1 x_{i,j} + b_2 y_{i,j} + b_3 z_{i,j} \ , \\ u_z(i, j) &= c_0 + c_1 x_{i,j} + c_2 y_{i,j} + c_3 z_{i,j} \ , \end{aligned} \tag{4}$$

where i and j span from $-m$ to m for a set of grid points; ($x_{i,j}$, $y_{i,j}$, $z_{i,j}$) is the corresponding 3D point location; while (a, b, c) are the polynomial coefficients to be determined. A large m tends to eliminate local variations in strain, while a small m may not sufficiently smooth strain estimation in certain regions. In this work, m was empirically chosen to be 4 as a reasonable trade-off.

A set of discrete surface points and their displacement vectors generated from motion tracking, stereovision, and smoothing were used to calculate in-plane strains. The computations were facilitated by triangulating the points into a set of finite elements to approximate element strains using linear shape functions. Because the brain surface is not coplanar, a single 2D coordinate system is not sufficient for representation. Instead, local 2D coordinate systems were constructed for each triangular element with the *x*-axis parallel to one edge, and the in-plane nodal displacements (u_i, u_j) were transformed into these coordinates (Fig. 1).

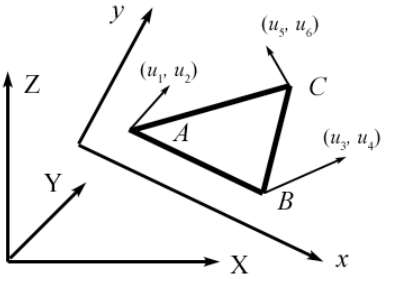


Fig. 1. Schematic of transforming a triangular element in 3D coordinates (X, Y, Z) into a 2D local coordinate system (*x*, *y*) with the *x*-axis parallel to one edge

The strain of each element, $\boldsymbol{\varepsilon}$ (defined as $\{\varepsilon_x\ \varepsilon_y\ 2\gamma_{xy}\}^T$), was determined from:

$$\boldsymbol{\varepsilon} = \begin{Bmatrix} \partial u_x/\partial x \\ \partial u_y/\partial y \\ \partial u_x/\partial y + \partial u_y/\partial x \end{Bmatrix} = \begin{bmatrix} \partial/\partial x & 0 \\ 0 & \partial/\partial y \\ \partial/\partial y & \partial/\partial x \end{bmatrix} \times \begin{bmatrix} N_1 & 0 & N_3 & 0 & N_5 & 0 \\ 0 & N_2 & 0 & N_4 & 0 & N_6 \end{bmatrix} \times \{\mathbf{u}\}\ , \tag{5}$$

where N_i (i=1:6) are the element shape functions for the six in-plane displacement components (Fig. 1), and $\{\mathbf{u}\}$ are the nodal displacements [8]. Since the strain components depend on choice of coordinate system, we evaluated system-invariant in-plane *principal* strains by forming a strain tensor for each triangular element by re-arranging the strain components (assuming an infinitesimal strain theory):

$$\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_x & \gamma_{xy} \\ \gamma_{xy} & \varepsilon_y \end{bmatrix}. \tag{6}$$

The maximum and minimum components of the principal strain (denoted as ε^1 and ε^2, respectively) and their directions were then obtained by computing the Eigen values and Eigen vectors of the strain tensor [9].

2.4 Simulated Displacement Data and Clinical Cases

To verify algorithmic implementation for computing principal strains, two simulated displacement fields were generated where each grid point, (*x*, *y*), had a displacement, (u_x, u_y), governed by Eqns. 7 (for simple stretch) or 8 (for simple shear):

$$u_x = x/20 \text{ , and } u_y = 0 \text{ ,} \tag{7}$$

$$u_x = y/20 \text{ , and } u_y = 0 \text{ .} \tag{8}$$

The numerically computed principal strains were compared with their ground-truth counterparts ($\varepsilon^{1,2}$=(0.05,0) and $\varepsilon^{1,2}$=(0.025,–0.025) for stretch and shear, respectively).

For clinical data, stereo image sequences captured from three patients (an 18 year-old male with epilepsy, a 65 year-old male with glioblastoma, and a 61 year-old female with meningioma, respectively) after dural opening were used to evaluate the dynamic surface displacement and strain. Patient arterial blood pressure synchronized with stereo image acquisition was also recorded (at 1 kHz). For illustration, the first stereo image pair in each case was selected as the "baseline" from which cortical surface deformation was estimated by comparing with subsequent stereo image acquisitions. All data analyses were performed on a Linux computer (2.6 GHz, 8 GB RAM) using MATLAB (R2010b, The Mathworks, Natick, MA).

3 Results

The computed principal strains were plotted at the centroid of each element for the two simulated displacement fields (Fig. 2). Both magnitudes and orientations of the computed principal strains matched exactly with their ground-truth counterparts, suggesting a correct implementation of the algorithm for calculating principal strains.

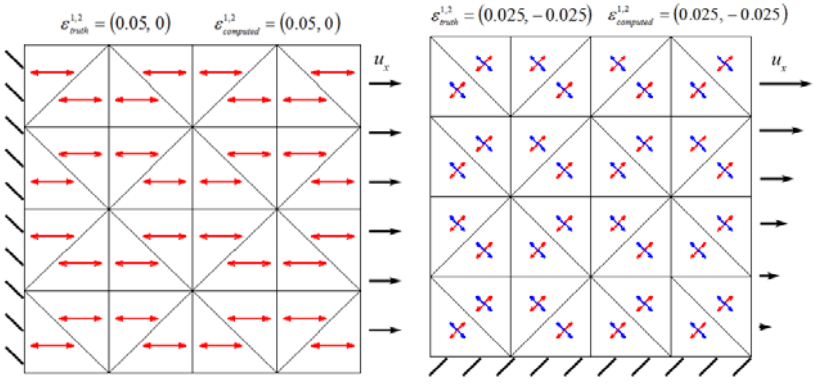


Fig. 2. The computed ε^1 (red) and ε^2 (blue) for the two simulated displacement fields (left: simple stretch in the *x* direction; right: simple shear)

Using clinical data, the rectified image corresponding to the largest relative displacement (on average) was overlaid on the baseline for patient 1 (Fig. 3a), where relative movements in two representative regions are shown (insets). The corresponding composite distribution of displacement components in the *x* and *y* directions was generated on the same rectified baseline image (Fig. 3b). A heterogeneous distribution of displacement components was evident that appears to correlate spatially with cortical areas defined by major vessels. Artifact due to

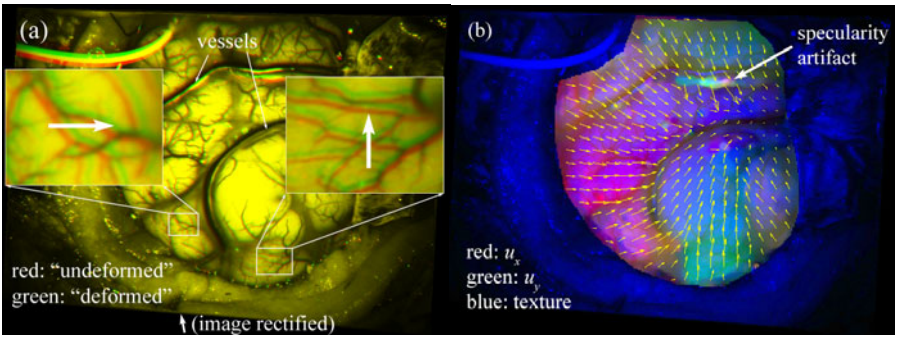


Fig. 3. A rectified image corresponding to the maximum relative displacement on average (green) overlaid on the baseline (red; a), showing distinctive motion patterns in two representative regions (insets; dominant motion represented by thick arrows). The resulting distribution of displacement components in the cortical area is also shown (b; displacement vectors are shown at one of every 9 points and are magnified by 10 times).

specularity (arrow in Fig. 3b) was confined locally, demonstrating the robustness of the motion-tracking algorithm in estimating the *overall* displacement distribution.

Cortical surface principal strains were computed by first transforming the locally smoothed displacements into 3D space through stereopsis. Representative distributions of ε^{I} for patients 1 (Fig. 4a) and 3 (Fig. 4b) are overlaid on their corresponding cortical surfaces reconstructed in 3D. For all patients, the largest strains occurred in areas near the craniotomy rim (up to 0.18; patient 2 not shown). The majority of ε^{I} in this region was perpendicular to the craniotomy boundary (Fig. 4), indicating relative stretching along this direction. The maximum average ε^{I} across the cortical area for the three patients was 0.067, 0.038, and 0.047, respectively.

Typical average displacement and ε^{I} as a function of time are shown for patient 3 in Fig. 5, clearly demonstrating that the cortical surface motion was in concert with blood pressure pulsation at a frequency of approximately 1 Hz. The peak values of displacement and ε^{I} as well as the normalized arterial blood pressure also varied in time, likely due to patient respiration (at a frequency of approximately 0.125 Hz).

4 Discussion and Conclusion

An accurate and comprehensive understanding of the state of soft tissue strain is important in determining tissue physiological and biomechanical properties that could provide new information for patient care and/or biomechanical modeling. Using stereovision, we have developed a completely noninvasive approach to estimate cortical surface strain from motion tracking. Our technique is based on spatially smooth full-field displacements that allow strain to be estimated over a complete 3D tissue surface determined from stereovision rather than from a single camera that is typically used in digital image correlation (DIC) which is popular in experimental mechanics [7]. In addition, the high-resolution images also allow a high sensitivity in displacement tracking (within 0.05–0.1 mm), which is important for high precision in

strain estimation. The sensitivity of displacement tracking and precision in strain estimation can be enhanced further by either increasing the camera image resolution or decreasing the field of view (e.g., through image magnification).

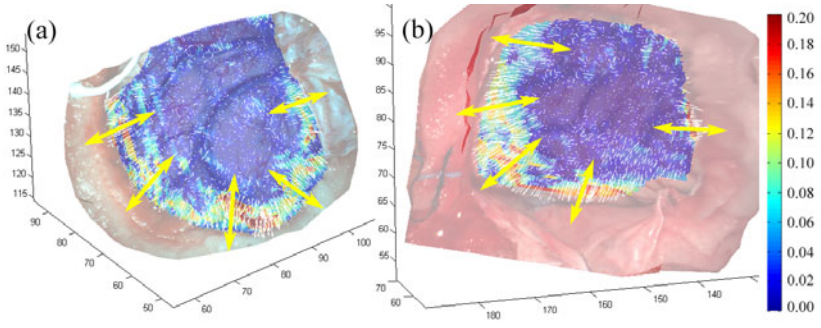


Fig. 4. Representative distributions of ε^1 for patient 1 (a; 0.067 on average) and 3 (b; 0.047 on average). Directions of ε^1 are shown at every 3rd point (magnified by 10 times), and a representative few are magnified further to enhance visualization (thick arrows).

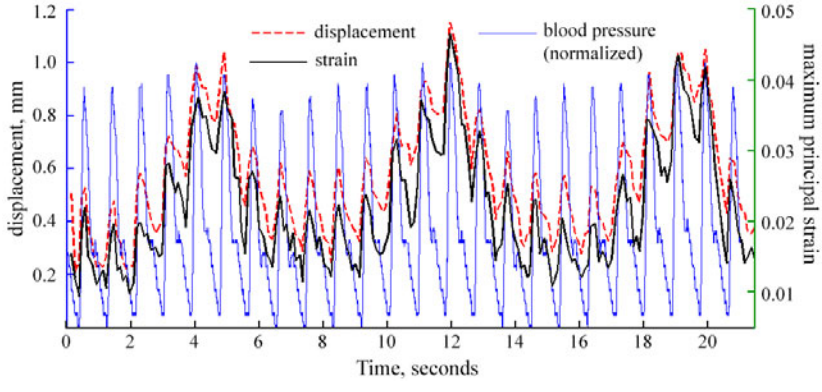


Fig. 5. Typical average displacement and ε^1 as a function of time for patient 3, demonstrating that cortical surface motion was in concert with blood pressure pulsation (normalized for visualization) at approximately 1 Hz. Stereo images were captured at a frequency of ~9.3 fps.

The application of the technique was demonstrated in three surgical cases where cortical surface motion due to arterial blood pressure pulsation was evaluated. Local displacement smoothing was first performed prior to differentiation to reduce noise. In addition, in-plane principal strains were calculated to characterize the 3D surface strain because they are invariant to the coordinate system selected. Results suggest that the dominant cortical surface strain occurred near the craniotomy rim (ε^1 up to ~18%), likely because of brain-skull friction when the parenchyma distended in this vicinity. When the brain otherwise sagged resulting in brain-skull separation (e.g., as a result of gravity), the stretching near the craniotomy boundary was not evident (e.g., see upper boundaries in Fig. 4a and b). While an accuracy evaluation is nearly

impossible because of the difficulties in determining ground-truth under clinical conditions, the dynamic evolution of displacement and strain over time with the synchronized blood pressure pulsation (Fig. 5) suggests our technique estimates strain with high fidelity (although the values are relative because they are based on a baseline that does not necessarily correspond to the true zero-state of strain). Nonetheless, our studies would benefit from additional quantitative assessments conducted in more controlled simulated and experimental environments in the future. Because of the relatively large strains especially around the craniotomy rim, finite strain theory which is applicable for arbitrarily large rotations and strains may also be more appropriate, and will be incorporated in the future (difference of ~10% at the 0.2 strain level is expected relative to the infinitesimal strain theory used in this study).

In summary, we have demonstrated the application of a completely noninvasive technique to estimate cortical surface strain from motion tracking during open cranial neurosurgery. The technique provides sensitive detection of surface displacement and tissue strain under physiological conditions as long as the surface is sufficiently rich in tracking features. With this technique, it may be possible to correlate the level of surface strain with tissue compliance [3], which is an important clinical indicator; or to provide tissue deformation feedback during surgical operations (e.g., tracking tissue strain around a retractor blade), which is important for patient care in practice.

Acknowledgement. Funding from the NIH grant R01 CA159324–01 is acknowledged.

References

1. Sun, H., Lunn, K.E., Farid, H., Wu, Z., Roberts, D.W., Hartov, A., Paulsen, K.D.: Stereopsis-Guided Brain Shift Compensation. IEEE Trans. Med. Imag. 24(8), 1039–1052 (2005)
2. Paul, P., Morandi, X., Jannin, P.: A Surface Registration Method for Quantification of Intraoperative Brain Deformations in Image-Guided Neurosurgery. IEEE Trans. Info. Tech. Biomed. 13(6), 976–983 (2009)
3. Girard, M.J.A., Downs, J.C., Burgoyne, C.F., Suh, J.K.F.: Experimental Surface Strain Mapping of Porcine Peripapillary Sclera Due to Elevations of Intraocular Pressure. J. of Biomech. Eng. 130, 041017–1 (2008)
4. Kiefer, M., Eymann, R.: Clinical Proof of the Importance of Compliance for Hydrocephalus Pathophysiology. Acta Neurochirurgica Supplementum 102, 115–118 (2009)
5. Liu, C.: Beyond Pixels: Exploring New Representations and Applications for Motion Analysis. Doctoral Thesis. Massachusetts Institute of Technology (May 2009)
6. Brox, T., Bruhn, A., Papenberg, N., Weickert, J.: High accuracy optical flow estimation based on a theory for warping. In: Pajdla, T., Matas, J(G.) (eds.) ECCV 2004. LNCS, vol. 3024, pp. 25–36. Springer, Heidelberg (2004)
7. Pan, B., Qian, K., Xie, H., Asundi, A.: Two-dimensional digital image correlation for in-plane displacement and strain measurement: a review. Meas. Sci. Technol. 20, 062001 (2009)
8. Zienkiewicz, O.C., Taylor, R.L., Zhu, J.Z.: The Finite Element Method: Its Basis and Fundamentals, 6th edn. Elsevier Butterworth-Heinemann, Oxford (2005)
9. Gould, P.L.: Introduction to Linear Elasticity, 2nd edn. Springer, Heidelberg (1993)

A Comparative Study of Physiological Models on Cardiac Deformation Recovery: Effects of Biomechanical Constraints

Ken C.L. Wong[1,2], Linwei Wang[1], Huafeng Liu[3,1], and Pengcheng Shi[1]

[1] Computational Biomedicine Laboratory, Rochester Institute of Technology, Rochester, USA
{linwei.wang,pengcheng.shi}@rit.edu
[2] ASCLEPIOS Research Project, INRIA Sophia Antipolis, Sophia Antipolis, France
chun-lok.wong@inria.fr
[3] State Key Laboratory of Modern Optical Instrumentation,
Zhejiang University, Hangzhou, China
liuhf@zju.edu.cn

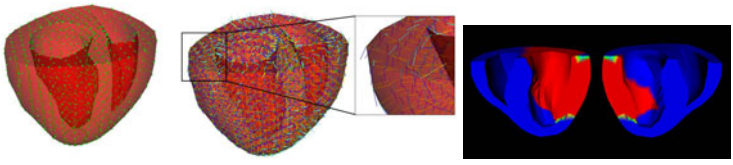
Abstract. Cardiac deformation recovery is to recover quantitative subject-specific myocardial deformation from imaging data. In the last decade, cardiac physiological models derived from anatomy, biomechanics, and cardiac electrophysiology have become increasingly popular in constraining the recovery problems because of their physiological meaningfulness. Although physiological models with various electrical and biomechanical components have been adopted by different frameworks and have exhibited promising results, these models have not been systematically compared under the same recovery framework, input data, and experimental setups. As different models comprise varying physiological plausibilities and complexities, comparisons under the same settings can aid choosing the proper models for specific goals and available resources. In this paper, under a state-space filtering framework for statistically optimal couplings between models and image data, we compare the performances of six different cardiac physiological models with different biomechanical constraints. Experiments were performed on synthetic data for quantitative comparisons, and on clinical data for their capabilities in identifying pathological situations.

1 Introduction

Cardiac deformation recovery is to recover quantitative subject-specific myocardial deformation from measurements. For noninvasive recoveries, medical images such as magnetic resonance images (MRI) provide an excellent source of *in vivo* anatomical and motion information. Nevertheless, as these measurements are sparse, projective, or noisy, and cannot directly provide quantitative cardiac deformation, *a priori* models are necessary to constrain the inverse problems for unique and meaningful results.

To provide physiologically meaningful constraints for deformation recovery from cardiac images, constraining models of different physiological plausibilities have started to be utilized in the last decade. In [1], a linear and isotropic biomechanical model (BM model) was used for statistical joint estimation of cardiac deformation and material properties. In [2], apart from the linear and transversely isotropic BM model with

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 420–427, 2011.



(a) (b) (c)

Fig. 1. Synthetic data. (a) Heart represented as nodes bounded by surfaces. (b) Tissue structure indicated by local coordinate systems (f,s,n: fiber, sheet, sheet normal: blue, yellow, cyan). (c) Infarcted regions shown in red.

fiber orientations considered, contraction forces were also simulated to provide an active deformation prior. In [3], a hyperelastic and orthotropic BM model was adopted with a nonlinear state-space filtering framework. The constraining models have become more physiologically plausible and complicated, from passive models with linear and isotropic materials to active models with hyperelastic and orthotropic materials.

These frameworks show promising results, however, as they were developed by different researchers at different times with different implementation difficulties, the models were not compared under the same recovery framework, input data, and experimental setups. As different models comprise varying physiological plausibilities and complexities, comparisons under the same settings can aid choosing the proper models for specific goals and available resources, so that unnecessary theoretical and implementation difficulties could be avoided. In view of this, we present here a comparative study of physiological models for cardiac deformation recovery from medical images using the recently developed framework in [3]. A nonlinearity-preserving state-space filtering framework based on unscented Kalman filter was used to couple physiological models with imaging data. Under this framework, the performances of six cardiac physiological models comprising different biomechanical material properties and force components were compared with the same input data and experimental settings. Experiments were performed on synthetic data for quantitative comparisons, and on clinical data for their capabilities in identifying pathological situations.

2 Cardiac Physiological Models

2.1 Cardiac System Dynamics

The heart is typically represented as a set of nodes bounded by the heart surfaces (Fig. 1(a)). The matrix representation of the cardiac system dynamics is given as:

$$\mathbf{M}\ddot{\mathbf{U}} + \mathbf{C}\dot{\mathbf{U}} + \mathbf{K}\Delta\mathbf{U} = \mathbf{F} \tag{1}$$

where $\mathbf{M}$, $\mathbf{C}$, and $\mathbf{K}$ are the mass, damping, and stiffness matrices respectively. $\ddot{\mathbf{U}}$, $\dot{\mathbf{U}}$, and $\Delta\mathbf{U}$ comprise the respective nodal accelerations, velocities, and incremental displacements. $\mathbf{F}$ contains nodal external forces applied to the system. Different constraining models comprise different material properties in $\mathbf{K}$ and external forces in $\mathbf{F}$.

Table 1. Physiological models compared in this study

Models	Stress-strain relation		Tissue structure		External forces	
	Linear	Nonlinear	Isotropic	Anisotropic	Passive	Active
LI-PBM	X		X		X	
LI-APM	X		X			X
LTI-PBM	X			X	X	
LTI-APM	X			X		X
HO-PBM		X		X	X	
HO-APM		X		X		X

2.2 Physiological Models for Comparisons

Physiological models with different material properties and external forces were tested. For the material properties in $\mathbf{K}$, hyperelastic and orthotropic (HO), linear and transversely isotropic (LTI), and linear and isotropic (LI) material models were tested. For the external forces in $\mathbf{F}$, the passive biomechanical model-based (PBM-based) recoveries with only image-derived forces, and the active physiological model-based (APM-based) recoveries with also simulated active contraction were tested. As the LI, LTI, or HO BM models can be used with either the PBM-based or the APM-based recoveries, a total of six models were studied (Table 1).

Material Properties. Different material properties associate with different assumptions of material nonlinearity and anisotropy of the myocardial tissue.

Hyperelastic and Orthotropic (HO) Material Model. Cardiac anatomy and biomechanics have shown that the myocardial tissue should be modeled as hyperelastic and orthotropic [4]. The HO model in [5] was used in the study, with strain energy function:

$$\Psi(\boldsymbol{\epsilon}) = \kappa(J \ln J - J + 1) + \frac{1}{2} a_{\mathrm{BM}}(e^{Q} - 1) \tag{2}$$

where

$$\begin{aligned} Q =& b_{ff}\overline{\epsilon}_{ff}^2 + b_{ss}\overline{\epsilon}_{ss}^2 + b_{nn}\overline{\epsilon}_{nn}^2 \\ &+ b_{fs}\left(\overline{\epsilon}_{fs}^2 + \overline{\epsilon}_{sf}^2\right) + b_{fn}\left(\overline{\epsilon}_{fn}^2 + \overline{\epsilon}_{nf}^2\right) + b_{sn}\left(\overline{\epsilon}_{sn}^2 + \overline{\epsilon}_{ns}^2\right) \end{aligned} \tag{3}$$

with J the determinant of deformation gradient, and κ the penalty factor for tissue incompressibility. $\overline{\epsilon}_{ij}$ are the isovolumetric components of the Green-Lagrange strain tensor $\boldsymbol{\epsilon}$. a_{BM} and b_{ij} are the material constants. The f-s-n coordinate system represents the fibrous-sheet structure (Fig. 1(b)).

Linear and Transversely Isotropic (LTI) Material Model. This model has been typically used for image-based cardiac deformation recovery [2]. The model is characterized by four material constants, E_f, E_{cf}, ν_f, ν_{cf}, which are the Young's moduli and Poisson's ratios along and across the fiber respectively.

Linear and Isotropic (LI) Material Model. This model was used in the earlier approaches [1]. The model is characterized by two material constants, with Young's modulus $E = E_f = E_{cf}$ and Poisson's ratio $\nu = \nu_f = \nu_{cf}$.

External Force Components. The force vector $\mathbf{F}$ in (1) drives the deformation, which can be provided by enforcing displacements of salient cardiac features derived from images [1], or by active contraction simulated using electrophysiological models [2,3].

Passive Biomechanical Model-Based (PBM-Based) Recoveries. The BM models are used without active contraction stresses. In consequence, the deformations are introduced passively by the image information through the filtering process in Section 3.

Active Physiological Model-Based (APM-Based) Recoveries. The BM models are used with simulated active contraction stresses. A two-variable diffusion-reaction system has been adopted to model the action potential propagation [6], which is further transformed into active stresses through the electromechanical coupling model in [2].

3 Cardiac Deformation Recovery with Statistically Optimal Model-Measurement Coupling

The cardiac physiological models provide physiological constraints for cardiac deformation recovery. Nevertheless, as the models are not subject-specific, and the measurements extracted from images are not perfect, a nonlinear state-space filtering framework is required to couple the measurements with the models according to their own merits.

To utilize state-space filtering, the stochastic state-space equations:

$$\mathbf{x}(k) = f\left(\mathbf{x}(k-1)\right) + \boldsymbol{\omega}(k-1) \tag{4}$$

$$\mathbf{y}(k) = h\left(\mathbf{x}(k)\right) + \boldsymbol{\nu}(k) \tag{5}$$

need to be defined. In (4), $\mathbf{x}$ is the state vector to be estimated with model uncertainties $\boldsymbol{\omega}$, and f projects $\mathbf{x}$ from time step $k-1$ to k. In (5), $\mathbf{y}$ is the measurement vector with measurement errors $\boldsymbol{\nu}$, and h relates the state to the measurements.

For cardiac deformation recovery, $\mathbf{x}$ is the nodal displacement vector $\mathbf{U}$. Using the relation $\mathbf{U}(k) = \mathbf{U}(k-1) + \Delta\mathbf{U}$, we have the concrete form of (4) as:

$$\mathbf{U}(k) = f\left(\mathbf{U}(k-1)\right) + \boldsymbol{\omega}(k-1) = \mathbf{U}(k-1) + \Delta\mathbf{U} + \boldsymbol{\omega}(k-1) \tag{6}$$

where $\Delta\mathbf{U}$ is obtained by solving (1), in which $\mathbf{K}$ depends nonlinearly on $\mathbf{U}$.

The concrete form of (5) is given as:

$$\mathbf{y}(k) = \mathbf{H}\mathbf{x}(k) + \boldsymbol{\nu}(k) \tag{7}$$

with $\mathbf{y}$ the nodal displacements of salient cardiac features extracted from images through motion tracking, and $\mathbf{H}$ the measurement matrix relating $\mathbf{U}$ to $\mathbf{y}$.

With (4,5), the model and measurements are connected together, and state-space filtering can be performed to obtain the optimal estimation. To preserve model nonlinearity and maintain computational feasibility, we utilize the unscented Kalman filter which comprises the advantages of Monte Carlo methods and Kalman filter updates [7].

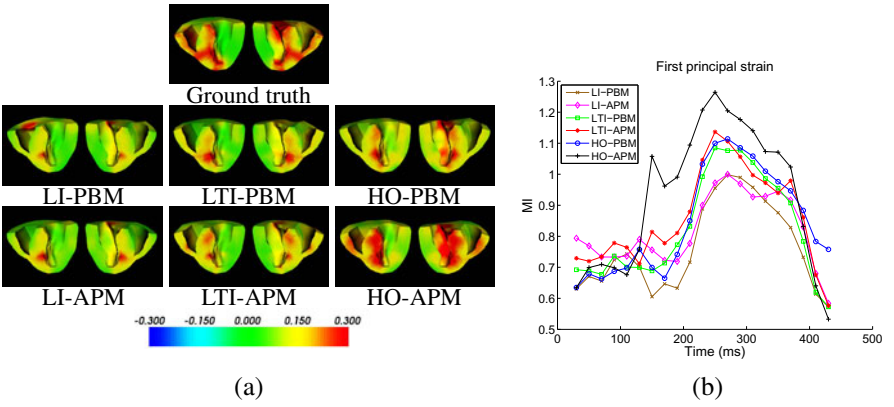


Fig. 2. Synthetic data. (a) First principal strains at the end of systole (250 ms). (b) Mutual information between strain patterns of the ground truth and the recovered deformations for the whole cardiac cycle.

4 Comparisons of Physiological Models

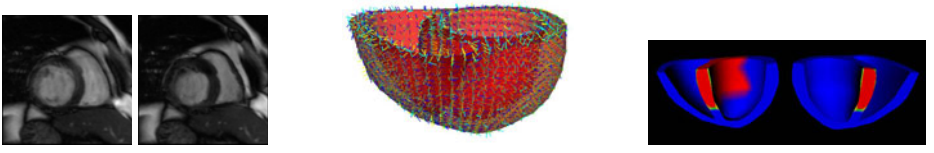
The model parameters were adopted from the literature. For the HO BM model, $a_{\mathrm{BM}} = 0.88$ kPa, $b_{ff} = 6$, $b_{ss} = 7$, $b_{nn} = 3$, $b_{fs} = 12$, $b_{fn} = 3$, $b_{sn} = 3$, and $\kappa = 5 \times 10^4$ [5]. For the LTI BM model, $E_f = 75$ kPa, $E_{cf} = 25$ kPa, $\nu_f = 0.4$, and $\nu_{cf} = 0.2$. For the LI BM model, $E = 75$ kPa and $\nu = 0.4$ [1].

4.1 Synthetic Data

Experimental Setups. The heart architecture from the University of Auckland was used to provide the cardiac geometry and tissue structure [8], and a heart representation of 1746 nodes was constructed (Fig. 1(a) and (b)). To verify the capabilities of locating diseased areas, some regions of the heart were set to be infarcted (Fig. 1(c)), where the electricity could not propagate through, and the stiffness was set to be three times as that in the normal regions. With this setting, the HO BM model in Section 2.2 was used to simulate a cardiac cycle of 450 ms as the ground truth. The nodal displacements on the heart boundaries were extracted and noises of 10dB signal-to-noise ratio were added, which were treated as the measurement inputs from medical images.

Results and Discussion. Fig. 2(a) shows the first principal strain maps for visual comparisons, as there is a strong inverse relation between the first principal strain and the extent of myocardial infarction [9]. There is almost no deformation in the infarcted regions in the ground truth. This observation also appears in the recovered strains of different models, but different models possess different similarities to the ground truth. To quantify the similarities of patterns rather than the absolute values, mutual information comparison is used (Fig. 2(b)). Higher mutual information implies higher similarity.

The strain maps and mutual information show that the PBM-based and APM-based recoveries behave similarly, as the measurements on the heart surfaces provided the displacement boundary conditions of the deformations. When the boundary conditions of



(a) (b) (c)

Fig. 3. Patient data. (a) MRI at end of diastole and end of systole. (b) Cardiac geometry segmented from the first frame of the MRI sequence, with mapped tissue structure (f,s,n: blue, yellow, cyan). (c) Infarcted segments (3, 4, 9, 10) shown in red.

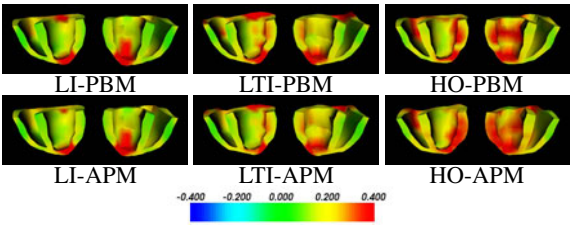


Fig. 4. Patient data. First principal strains at the end of systole.

most surfaces were given, the effects of the active stresses decreased. Such similarities are more obvious for the LI and LTI models, as the active forces were calibrated using the HO-APM model and only caused relatively small deformations when applied to the linear BM models. Therefore, the active forces did not contribute much to the recovery.

On the other hand, the differences between different BM models are more obvious. The differences between the LI and the LTI models are very clear, especially around the end of systole. As the only differences between these two linear BM models are the Young's modulus E_{cf} and the Poisson's ratio ν_{cf}, this shows the importance of using anisotropic BM model. For the HO models, the HO-APM model has the highest MI, but the HO-PBM model behaves very similar to the LTI-models. Therefore, HO models need to be used with proper active contraction stresses to gain the benefits.

4.2 Patient Data

Experimental Setups. The data sets are available in [10], for case 1 and case 2 from two patients with acute myocardial infarction, with the infarcted regions identified by experts. Case 1 contains a human short-axis MRI sequence of 19 frames in one cardiac cycle (52.5 ms/frame), with 12 slices/frame, 8 mm inter-slice spacing, and in-plane resolution 1.32 mm/pixel. Case 2 contains a human short-axis MRI sequence of 16 frames in one cardiac cycle (50 ms/frame), with 13 slices/frame, 8 mm inter-slice spacing, and in-plane resolution 1.32 mm/pixel. In this paper, only the figures of case 2 are shown for illustration. Fig. 3(a) shows the MRI. Segmentations were performed to obtain the heart geometries at the end of diastole, and the fibrous-sheet structures were mapped from the Auckland heart architecture using nonrigid registration (Fig. 3(b)). Fig. 3(c) shows the infarcted regions identified by experts from contrast-enhanced MRI with the

Table 2. Patient data. Segment numbers at different levels sorted by the mean of the first principal strains of each segment at the end of systole, in ascending order. The numbers of the infarcted segments are highlighted in red. The optimal cutoff segments for identifying infarcted segments are indicated through double vertical lines (‖).

Case 1	Basal	Mid	Apical
LI-PBM	3 2 4 1 ‖ 6 5	10 9 8 ‖ 7 11 12	14 15 13 ‖ 16
LI-APM	3 4 2 1 ‖ 6 5	10 9 8 ‖ 11 7 12	14 15 13 ‖ 16
LTI-PBM	3 4 5 2 1 ‖ 6	9 10 8 ‖ 11 7 12	14 13 15 ‖ 16
LTI-APM	3 4 2 ‖ 5 1 6	9 8 ‖ 10 11 7 12	14 13 15 ‖ 16
HO-PBM	3 2 4 1 ‖ 5 6	9 8 ‖ 10 12 7 11	14 13 15 ‖ 16
HO-APM	3 2 4 1 ‖ 5 6	8 9 ‖ 10 12 7 11	14 13 15 ‖ 16

Case 2	Basal	Mid	Apical
LI-PBM	3 ‖ 6 5 1 2 4	9 12 10 ‖ 8 7 11	15 14 16 13
LI-APM	3 ‖ 6 5 1 4 2	9 12 10 ‖ 8 11 7	15 14 16 13
LTI-PBM	3 ‖ 2 1 6 5 4	9 10 ‖ 12 7 8 11	15 14 16 13
LTI-APM	3 ‖ 2 5 1 6 4	9 10 ‖ 12 8 7 11	15 14 16 13
HO-PBM	3 1 4 ‖ 6 5 2	10 9 ‖ 11 12 8 7	15 14 16 13
HO-APM	1 3 4 ‖ 5 6 2	10 9 ‖ 12 7 11 8	15 14 16 13

Table 3. Patient data. Optimal sensitivities (SEN) and specificities (SPEC) for identifying infarcted segments.

Case 1	Basal	Mid	Apical
	SEN / SPEC	SEN / SPEC	SEN / SPEC
LI-PBM	100% / 67%	100% / 75%	100% / 100%
LI-APM	100% / 67%	100% / 75%	100% / 100%
LTI-PBM	100% / 33%	100% / 75%	100% / 100%
LTI-APM	67% / 67%	100% / 100%	100% / 100%
HO-PBM	100% / 67%	100% / 100%	100% / 100%
HO-APM	100% / 67%	100% / 100%	100% / 100%

Case 2	Basal	Mid
	SEN / SPEC	SEN / SPEC
LI-PBM	50% / 100%	100% / 75%
LI-APM	50% / 100%	100% / 75%
LTI-PBM	50% / 100%	100% / 100%
LTI-APM	50% / 100%	100% / 100%
HO-PBM	100% / 75%	100% / 100%
HO-APM	100% / 75%	100% / 100%

segmentation and nomenclature suggested by the American Heart Association. The infarcted segments of case 1 are 1, 2, 3, 8, 9, 13, 14, 15, and of case 2 are 3, 4, 9, 10. As short-axis MRI cannot provide accurate motion at the apex, segment 17 was not considered in our experiments. To obtain the measurement inputs, a registration framework based on free-form deformation was used to extract the displacement field from the image sequences [11], and the displacements on the current heart boundaries were utilized. The infarcted regions were unknown in the recoveries.

Results and Discussion. The qualitative assessments are shown in Fig. 4. The strain patterns of all models show relatively small deformations in the infarcted regions. Similar to the synthetic data, the PBM-based and the APM-based models are relatively similar, and the differences between different BM models are more obvious. All explanations for the strain patterns of the synthetic data are applicable here.

For clinical applications, the capabilities of frameworks in locating diseased regions are of great interest, thus the results are compared in this aspect. We sorted the segments at different levels by the mean of the first principal strains of each segment at the end of systole, in ascending order (Table 2). From Table 2, we can calculate the various sensitivities and specificities of the frameworks with different cutoff segments, where sensitivities are the ratios between the correctly identified positives (infarcted segments) and the actual positives, and the specificities are the ratios between the correctly identified negatives (normal segments) and the actual negatives. The optimal sensitivities and specificities are shown in Table 3. As there is no infarction at the apical level for case 2, the corresponding optimal sensitivities and specificities are unavailable.

Table 2 shows again that the differences between PBM-based and APM-based models are small. Table 3 shows that the HO models have the best capabilities in both case 1 and 2. Comparing between the LTI and LI models, in case 1, the LTI models have

better capabilities than the LI models at the mid-ventricular level but worse capabilities at the basal level. In case 2, the LTI models have better capabilities at the mid-ventricular level than the LI models. Considering also the results in the synthetic experiments, the LTI models perform better than the LI models.

5 Conclusion

The results show that when image information is used, the active contraction stresses do not contribute much in the recovery. Therefore, if subject-specific active contraction stresses are unavailable but the image quality is good, the PBM-based models can be the proper choices. On the other hand, the anisotropies of the BM models have more impacts on the results, but the HO models do not necessarily provide much better results than the LTI models especially in verifying infarcted regions. Therefore, if the absolute strain values are unimportant, the use of LTI models might be enough.

References

1. Shi, P., Liu, H.: Stochastic finite element framework for simultaneous estimation of cardiac kinematic functions and material parameters. Medical Image Analysis 7, 445–464 (2003)
2. Sermesant, M., Delingette, H., Ayache, N.: An electromechanical model of the heart for image analysis and simulation. IEEE Transactions on Medical Imaging 25(5), 612–625 (2006)
3. Wong, K.C.L., Wang, L., Zhang, H., Shi, P.: Nonlinear cardiac deformation recovery from medical images. In: IEEE International Conference on Image Processing, pp. 2513–2516 (2009)
4. Glass, L., Hunter, P., McCulloch, A. (eds.): Theory of Heart: Biomechanics, Biophysics, and Nonlinear Dynamics of Cardiac Function. Springer, Heidelberg (1991)
5. Usyk, T.P., Mazhari, R., McCulloch, A.D.: Effect of laminar orthotropic myofiber architecture on regional stress and strain in the canine left ventricle. Journal of Elasticity 61, 143–164 (2000)
6. Aliev, R.R., Panfilov, A.V.: A simple two-variable model of cardiac excitation. Chaos, Solitons & Fractals 7(3), 293–301 (1996)
7. Julier, S.J., Uhlmann, J.K.: Unscented filtering and nonlinear estimation. Proceedings of the IEEE 92(3), 401–422 (2004)
8. Nash, M.: Mechanics and Material Properties of the Heart using an Anatomically Accurate Mathematical Model. PhD thesis, The University of Auckland (1998)
9. Tim Marcus, J., Götte, M.J.W., van Rossum, A.C., Kuijer, J.P.A., Heethaar, R.M., Axel, L., Visser, C.A.: Myocardial function in infarcted and remote regions early after infarction in man: assessment by magnetic resonance tagging and strain analysis. Magnetic Resonance in Medicine 38(5), 803–810 (1997)
10. PhysioNet/Computers in Cardiology challenge: electrocardiographic imaging of myocardial infarction (2007), http://www.physionet.org/challenge/2007/
11. Perperidis, D., Mohiaddin, R.H., Rueckert, D.: Spatio-temporal free-form registration of cardiac MR image sequences. Medical Image Analysis 9(5), 441–456 (2005)

Motion Tracking and Strain Map Computation for Quasi-Static Magnetic Resonance Elastography

Y.B. Fu[1,*], C.K. Chui[1], C.L. Teo[1], and E. Kobayashi[2]

[1] Department of Mechanical Engineering, National University of Singapore, Singapore
[2] Department of Precision Engineering, The University of Tokyo, Tokyo, Japan
g0800248@nus.edu.sg

Abstract. This paper presents a new imaging method for quasi-static magnetic resonance elastography (MRE). Tagged magnetic resonance (MR) imaging of human lower leg was acquired with probe indentation using a MR-compatible actuation system. Indentation force was recorded for soft tissue elasticity reconstruction. Motion tracking and strain map of human lower leg are calculated using a harmonic phase (HARP)-based method. Simulated tagged MR images were constructed and analyzed to validate the HARP-based method. Our results show that the proposed imaging method can be used to generate accurate motion distribution and strain maps of the targeted soft tissue.

Keywords: MR tagging, HARP, motion tracking, strain, soft tissue indentation.

1 Introduction

Elasticity is an important property for characterization of biological soft tissues. Local changes in mechanical properties of soft tissues may indicate the presence of tumors and other diseases [1]. In order to quantitatively study the soft tissue mechanical properties, Magnetic Resonance Elastography (MRE) has been developed in recent years. MRE can be categorized into dynamic MRE [2-4] and static or quasi-static MRE [5,6]. However, there are less literature about the quasi-static MRE due to the difficulty of motion tracking and strain calculation of soft tissue.

Various techniques, such as Ultrasound, Computed Tomography (CT), and Magnetic Resonance Imaging (MRI) have been used to image the spatial deformation of soft tissue. Tagged MRI is widely used for the quantification of motion and deformation of cardiac tissue due to its easy access and good performance. However, the applicability of Tagged MRI on other parts of human body has not been investigated. In this paper, we extended the use of tagged MRI technique to other biological soft tissue by building a MR compatible actuating device synchronized with simulated electrocardiogram (ECG) signal. Large indentation tests on human lower leg were conducted using this imaging method.

Template Matching [7-10], Active Geometry [11-13], Optical Flow [14-17] and Harmonic Phase (HARP) [18] are the main methods for tagged MR images processing. Template matching methods calculate the displacement by tracking the

* Corresponding author.

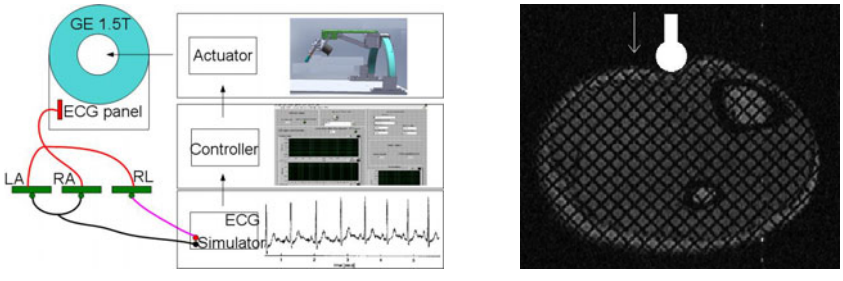
G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 428–435, 2011.

tag lines. Optical flow is advantageous over template matching in providing a dense estimation of the motion field instead of a sparse set of data located at the tag lines. However, the optical flow methods require a material point with constant brightness which is not available for tagged MR images due to tag fading. The HARP method proposed by Osman et al. [20] is widely spread in cardiac image processing [21,23]. However, the applicability of HARP analysis on soft tissue other than cardiac tissue has not been investigated. In this paper, the applicability and accuracy of HARP analysis for automatic motion tracking and strain calculation was investigated on tagged MR images of human lower leg.

2 Materials and Methods

2.1 MR Imaging

A MR-Compatible actuation system is developed for tagged MR imaging of soft tissue indentation (Fig. 1(a)). ECG signal is simulated and outputted to both the motor controller and MR scanner. MRI k-space data acquisition is synchronized with the simulated ECG signal. Specific k-space segment data is acquired and filled into corresponding images repeatedly after each ECG period. Twenty images are acquired over multiple ECG periods within time duration of 10-20s with GE SIGNA 1.5T MRI Scanner. Ultrasonic motor is controlled by microprocessor based controller interfaced with USB-6221 DAQ device. The control software and signal processing is implemented using LabView version 8.6. Motor is activated to rotate when *R* peak of simulated ECG is detected and stop after one full rotation. Indentation of the actuation device is synchronized with the simulated ECG signal. This ensures the deformation of imaged object to be consistent during each ECG period.



a b

Fig. 1. a: Overview of imaging system, b: Human lower leg indentation diagram

Indentation experiment was conducted on a volunteer's lower leg. Indentation force was recorded by a force sensor. The lower leg was indented at the top middle region of the leg using a sphere-shaped indenter with a diameter of 15mm (Fig. 1(b).). MR tags were generated by applying the tagging sequence immediately after the detection of the R-wave. Gradient-echo images were acquired during the entire imaging cycle to capture the displacement of the tags. Tagging period of 5mm is used.

Other imaging parameters are as follows: flip angle, 15°; tag orientation, 45°; echo time (TE), 1.704msec; repetition time (TR), 6.032msec; slice thickness, 15mm; spacing between slices, 15mm; field of view (FOV), 20×20 cm. The first and the seventh frames of the tagged images are shown in Fig. 1(b) and Fig. 4(a).

2.2 HARP Analysis

The spatial modulation of magnetization (SPAMM) sequence modulates the longitudinal magnetization in a sinusoidal manner, which produces an array of spectral peaks (harmonic peaks) in the Fourier domain (k-space). A circular band-pass filter (Fig. 4(b)) is often used to extract the off-center harmonic peaks. Harmonic phase image can be calculated by applying the inverse Fourier transform of a harmonic peak. Since the harmonic image is complex, it has both magnitude and phase at each pixel. The magnitude image reflects the anatomy of the lower leg and is used for segmentation of the leg in this study. The harmonic phase angle of the leg can be used to infer the tissue material property and remains invariant through the image sequences [19]. Displacement distribution and strain maps are calculated from these harmonic phase images.

Motion Tracking. The phase of the image at time t is given by

$$\emptyset(\boldsymbol{y}, t) = \boldsymbol{w}^T \boldsymbol{x}(\boldsymbol{y}, t) \tag{1}$$

where $\emptyset(\boldsymbol{y}, t)$ is the phase, $\boldsymbol{w}$ is the frequency vector of a harmonic spectral peak, $\boldsymbol{x}$ is the spatial position, $\boldsymbol{y}$ is the image coordinate [20]. At time *t*, the phase shift can be calculated as

$$\emptyset(\boldsymbol{y}, t) - \emptyset(\boldsymbol{y}, 0) = \boldsymbol{w}^T\big(\boldsymbol{x}(\boldsymbol{y}, t) - \boldsymbol{x}(\boldsymbol{y}, 0)\big) = \boldsymbol{w}^T \boldsymbol{u}(\boldsymbol{y}) \tag{2}$$

where $\boldsymbol{u}(\boldsymbol{y})$ is the displacement vector. Since the phase $\emptyset(\boldsymbol{y}, t)$ differs from the harmonic phase angle $a(\boldsymbol{y}, t)$ by a multiple of 2π,

$$\Delta a = wrap\big(a(\boldsymbol{y}, t) - a(\boldsymbol{y}, 0)\big) = wrap\big(\boldsymbol{w}^T \boldsymbol{u}(\boldsymbol{y})\big) \tag{3}$$

where the nonlinear wrapping function is given by

$$wrap(\emptyset) = mod(\emptyset + \pi, 2\pi) - \pi. \tag{4}$$

If $|\boldsymbol{w}^T \boldsymbol{u}(\boldsymbol{y})| < \pi$, $\boldsymbol{u}(\boldsymbol{y})$ can be calculated by

$$\boldsymbol{u}(\boldsymbol{y}) = (\boldsymbol{w}^T)^{-1} \Delta a. \tag{5}$$

If $|\boldsymbol{w}^T \boldsymbol{u}(\boldsymbol{y})| > \pi$, Δa is always wrapped. For soft tissue indentation, it is reasonable to assume that the displacement varies smoothly within the soft tissue. Given this assumption, $\boldsymbol{w}^T \boldsymbol{u}(\boldsymbol{y})$can be calculated by unwrapping Δa starting from a pixel where $|\boldsymbol{w}^T \boldsymbol{u}(\boldsymbol{y})| < \pi$ is satisfied. In this study, a quality guided phase unwrapping method is used.

Strain Calculation. Once the displacement field $\boldsymbol{u}(\boldsymbol{y})$ is calculated, the deformation gradient tensor of the lower leg is given by

$$\boldsymbol{F} = \nabla_{\mathrm{y}} \boldsymbol{u}(\boldsymbol{y}) + \boldsymbol{I}. \tag{6}$$

Therefore the 2D Lagrangian strain tensor can be calculated as

$$\boldsymbol{E} = \frac{1}{2}(\boldsymbol{F}^T\boldsymbol{F} - \boldsymbol{I}). \tag{7}$$

However, this method does not perform well in the presence of noise since the phase unwrapping procedure is needed to calculate $\boldsymbol{u}(\boldsymbol{y})$. The strain map is calculated using

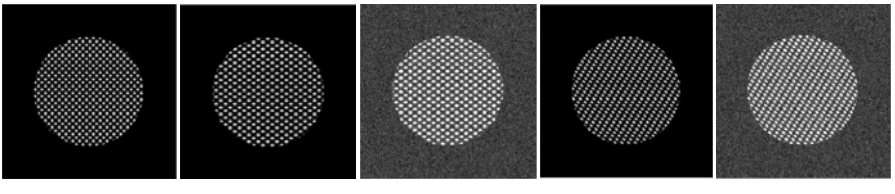
$$\boldsymbol{E}(\boldsymbol{y}) = \frac{1}{2}\left(\boldsymbol{H}^T\boldsymbol{w}\left(\frac{\partial\emptyset}{\partial\boldsymbol{y}}\right)^{-T} \cdot \left(\frac{\partial\emptyset}{\partial\boldsymbol{y}}\right)^{-1}\boldsymbol{w}^{\mathrm{T}}\boldsymbol{H} - \boldsymbol{I}\right) \tag{8}$$

where $\boldsymbol{E}(\boldsymbol{y})$ is the 2D Lagrangian strain, $\boldsymbol{H}$ matrices related with the tagging and imaging plane [20,21].

3 Results

3.1 Validation Using Simulated Tagging Images

Simulated tagging images were used to validate the strain calculation. The simulated tagging images were constructed by applying sinusoidal modulation to a digital circular image. A circular band-pass filter was used to extract the off-center harmonic peak. The radius of the filter was chosen at one third of the tagging frequency. Stretch strains ranging from 0.1 to 0.4 and shear strains ranging from 0.05 to 0.2 were constructed and processed (Fig. 2(a-e)). 100% Gaussian white noise was added to test the sensitivity of the algorithm to noise. The calculated strains show good agreement with the true values. The error of E_{xx} is calculated to be within ±8.07% and ±7.21% for images with and without noise respectively. The error of E_{xy} is within ±14.95% and ±13.65% for images with and without noise respectively. The strain calculation is accurate even in the presence of noise (Fig. 3(a-d)).



a b c d e

Fig. 2. Simulated images. a: Reference image. b: Deformed image of $E_{xx} = 0.4$ without noise. c: Deformed image of $E_{xx} = 0.4$ with noise. d: Deformed image of $E_{xy} = 0.2$ without noise. e: Deformed image of $E_{xy} = 0.2$ with noise.

3.2 Imaging of Human Lower Leg

The bones of the lower leg were segmented and excluded from the strain map computation. We processed only soft tissue of the leg. A circular band-pass filter with radius of 1.56mm was used to extract the off-center harmonic peak. The calculated harmonic phase image of one direction is shown in Fig. 4(c). We observed from Fig.

4(a) and Fig. 4(c) that the tag lines correspond well with the wrapped harmonic phase lines. Since the lower leg was indented from the top and around the middle region of the leg, the left and right regions of the lower leg were expectedly having negative and positive displacement values respectively in the x direction (Fig. 5(a)). In the y direction, area located near the indenter has larger displacement (Fig. 5(b)). Since the lower leg was compressed, negative strain values were expected to dominate in both directions. In y direction, smaller strain values were shown in areas away from the indenter and behind the bones (Fig. 5(d)). Due to the presence of noise, strain variation was observed in the bottom left and top right regions.

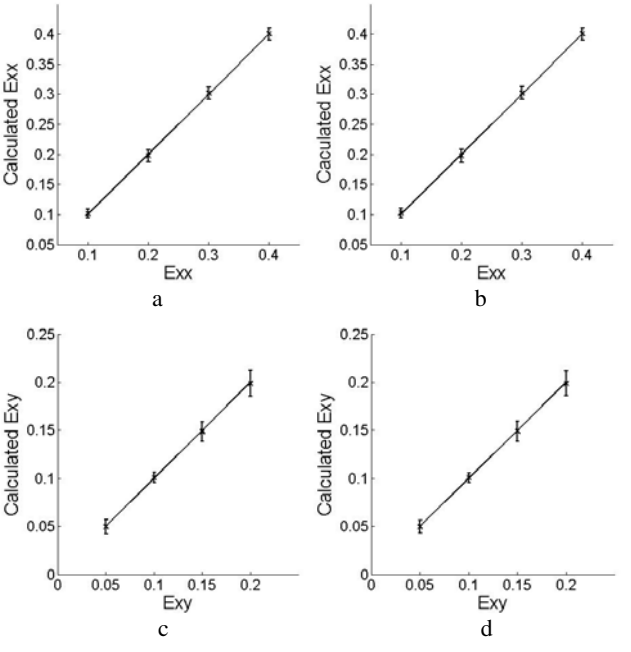


Fig. 3. Strain calculation with simulated images. a: Calculated strain of E_{xx} without noise. b: Calculated strain of E_{xx} with noise. c: Calculated strain of E_{xy} without noise. d: Calculated strain of E_{xy} with noise.

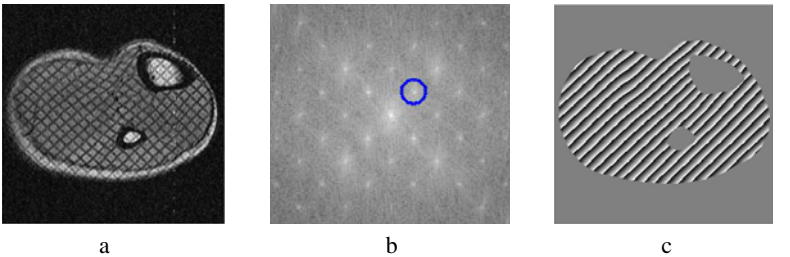


Fig. 4. a: Frame 7 of tagging image, b: Band-pass filter in Fourier domain, c: Wrapped harmonic phase image in one direction

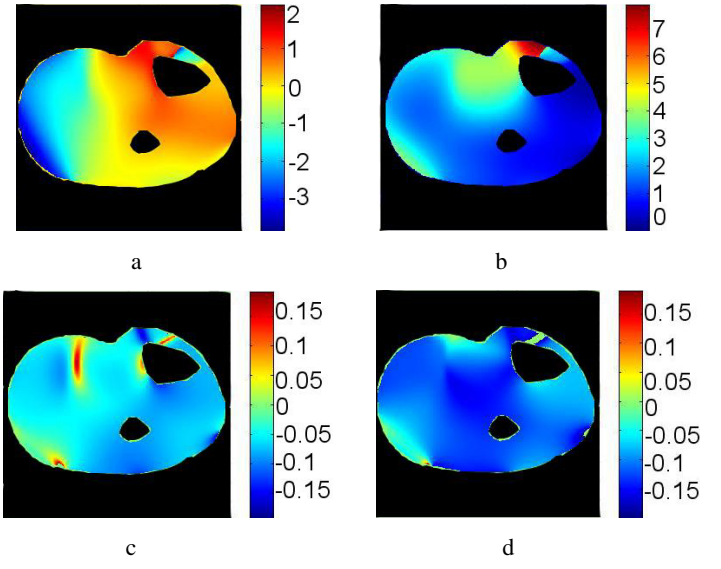


Fig. 5. a: Displacement in the x direction (unit: mm), b: Displacement in the y direction (unit: mm), c: Lagrangian strain in the x direction, d: Lagrangian strain in the y direction

4 Discussion and Conclusion

The proposed imaging system is a clinically viable image acquisition method for quasi-static MRE. To date, most MR strain imaging method for quasi-static MRE requires extensive programming of imaging sequence [5, 22] which can be challenging. In contrast, Tagged MRI is an imaging method widely equipped in almost every clinical MRI scanner. Using tagged MRI and the proposed method, strain imaging can be performed without imaging sequence programming. In this paper, tagging images of human lower leg indentation was obtained by building a MR-compatible actuator. Indentation force is recorded at the same time. Motion tracking and strain map of a human leg are calculated using the HARP-based method.

Motion tracking is performed by first measuring the phase shift of soft tissue during indentation. Displacement distribution maps were generated and analyzed. One limitation of the proposed motion tracking method is that it only applies to situation where smooth deformation presents. Validation of the HARP strain calculation is performed on a series of simulated tagged images. Strong agreement between the calculated and true strains is observed. The good noise reduction performance is due to the band-pass filter which also serves as a noise filter. The radius of the band-pass filter is important. Large radius may introduce unnecessary noise while small radius may suppress the actual deformation signal. Optimal radius selection is proposed in [20].

Although the current study was performed on human lower leg, our actuation system can be readily applied to other parts such as abdominal organs, breasts and foot. HARP analysis of images of these organs can be performed to obtain displacement distribution and strain maps. The elasticity distribution of these organs can then be calculated using various elasticity reconstruction methods.

Acknowledgement. This research is partially supported by NUS-JSPS Joint Research Project Grant.

References

1. Srinivasan, S., Krouskop, T., Ophir, J.: A quantitative comparison of modulus images obtained using nano indentation with strain elastograms. Ultrasound Med. Biol. 30(7), 899–918 (2004)
2. Ingolf, S., Johannes, B., Jurgen, B.: Analysis of wave patterns in MR elastography of skeletal muscle using coupled harmonic oscillator simulations. Magn. Reson. Imaging 20, 95–104 (2002)
3. Hamhaber, U., Sack, I., Papazoglou, S., Rump, J., Klatt, D., Braun, J.: Three-dimensional analysis of shear wave propagation observed by in vivo magnetic resonance elastography of the brain. Acta Biomater 3, 127–137 (2007)
4. Scott, A.K., Gregory, H.R., Kevin, J.G., Armando, M., Joel, P.F., Clifford, R.J., Richard, L.E.: Magnetic resonance elastography of the brain. Neuro. Image 39, 231–237 (2008)
5. Steele, D.D., Chenevert, T.L., Skovoroda, A.R., Emelianov, S.Y.: Three-dimensional static displacement, stimulated echo NMR elasticity imaging. Phys. Med. Biol. 45, 1633–1648 (2000)
6. Donald, B.P., Jonathan, B., Abbas, S., Justin, S.: Visualization and quantification of breast cancer biomechanical properties with magnetic resonance elastography, Phys. Med. Biol. 45, 1591–1610 (2000)
7. Atalar, E., McVeigh, E.: Optimization of tag thickness for measuring position with magnetic resonance imaging. IEEE Trans. Med. Imag. 13(1), 152–160 (1994)
8. Guttman, M.A., Prince, J.L., McVeigh, E.R.: Tag and contour detection in tagged MR images of the left ventricle. IEEE Trans. Med. Imaging 13(1), 74–88 (1994)
9. Kerwin, W.S., Prince, J.L.: Tracking MR tag surfaces using a spatiotemporal filter and interpolator. Int. J. Imag. Sys. Tech. 10(2), 128–142 (1999)
10. Deng, X., Thomas, S.D.: Combined tag tracking and strain reconstruction from tagged cardiac MR images without user-defined myocardial contours. J. Magn. Reson. Imaging 21(1), 12–22 (2005)
11. Kumar, S., Goldgof, D.: Automatic tracking of SPAMM grid and the estimation of deformation parameters from cardiac MR images. IEEE Trans. Med. Imag. 13, 122–132 (1993)
12. Amini, A.A., Chen, Y., Curwen, R.W., Mani, V., Sun, J.: Coupled Bsnake grids and constrained thin-plate splines for analysis of 2-D tissue deformations from tagged MRI. IEEE Trans. Med. Imag. 17(3), 344–356 (1998)
13. Qian, Z., Huang, X., Metaxas, D., Axel, L.: Robust segmentation of 4D cardiac MRI-tagged images via spatio-temporal propagation. In: Proc. SPIE Med. Imag., pp. 580–591 (2005)
14. Prince, J.L., McVeigh, E.R.: Motion estimation from tagged MR image sequences. IEEE Trans. Med. Imag. 11(2), 238–249 (1992)
15. Denney, J.T., Prince, J.L.: Optimal brightness functions for optical flow estimation of deformable motion. IEEE Trans. Image Proc. 3(2), 178–191 (1994)
16. Gupta, S.N., Prince, J.L.: On variable brightness optical flow for tagged MRI. In: Bizais, Y., Barillot, C., DiPaola, R. (eds.) Proc. Inf. Process. Med. Imag., pp. 323–334 (1995)
17. Haber, I., Kikinis, R., Westin, C.-F.: Phase-driven finite element model for spatio-temporal tracking in cardiac tagged MRI. In: Niessen, W.J., Viergever, M.A. (eds.) MICCAI 2001. LNCS, vol. 2208, pp. 1332–1335. Springer, Heidelberg (2001)

18. Osman, N.F., Kerwin, W.S., McVeigh, E.R., Prince, J.L.: Cardiac motion tracking using cine harmonic phase (harp) magnetic resonance imaging. Magn. Reson. Med. 42, 1048–1060 (1999)
19. Zhang, S., Douglas, M., Yaroslavsky, L., Summers, R., Dilsizen, V., Fananapazir, L., Bacharach, S.: A Fourier based algorithm for tracking spamm tags in gated magnetic resonance cardiac images. Med. Phys. 23(8), 1359–1369 (1996)
20. Osman, N.F., McVeigh, E.R., Prince, J.L.: Imaging heart motion using harmonic phase MRI. IEEE Trans. Med. Imaging 19, 186–202 (2000)
21. Liu, W., Chen, J.J., Ji, S.B., Allen, S.J., Bayly, P.V., Wickline, S.A., Yu, X.: Harmonic Phase MR Tagging for Direct Quantification of Lagrangian Strain in Rat Hearts after Myocardial Infarction. Magn. Reson. Med. 52, 1282–1290 (2004)
22. Chenevert, T.L., Skovoroda, A.R., O'Donnell, M., Emelianov, S.Y.: Elasticity Reconstructive Imaging by Means of Stimulated Echo MRI. Magn. Reson. Med. 39, 482–490 (1998)
23. Pan, L., Prince, J.L., Lima, J.A., Osman, N.F.: Fast tracking of cardiac motion using 3D-HARP. IEEE Trans. Biomed. Eng. 52(8), 1425–1435 (2005)

Blood Flow Computation in Phase-Contrast MRI by Minimal Paths in Anisotropic Media

Michael Schwenke[1], Anja Hennemuth[1], Bernd Fischer[2,3], and Ola Friman[1]

[1]Fraunhofer MEVIS - Institute for Medical Image Computing, Bremen, Germany
[2]Fraunhofer MEVIS - Project Group Image Registration, Lübeck, Germany
[3]Institute of Mathematics and Image Computing - University of Lübeck, Germany
michael.schwenke@mevis.fraunhofer.de
http://www.mevis.fraunhofer.de

Abstract. In this paper, anisotropic Fast Marching is employed to compute blood flow trajectories as minimal paths in 3D phase-contrast MRI images. Uncertainty in the estimated blood flow vectors is incorporated in a tensor which is used as metric for the anisotropic Fast Marching. A flow connectivity distribution is computed simultaneously to the Fast Marching. Based on the connectivity distribution the most likely flow trajectories can be identified. Results are presented for several PC MRI data sets and the capability of the method to indicate uncertainty of the flow trajectories is shown.

Keywords: blood flow computation, uncertainty, minimal path, anisotropic Fast Marching

1 Introduction

Fast Marching (FM) methods have been widely used for segmentation and analysis of medical images [9]. The FM method is a numerical scheme for solving the so-called Eikonal differential equation, whose solution can be interpreted as the monotone evolution of an interface on a domain with a spatially varying speed function. A main application of FM methods is to find a minimal path that describes the fastest way to travel between two points considering the speed function. The minimal path can for example represent a segmentation contour or the centerline of a vessel [2,4]. A generalization of the original FM method is obtained by considering *anisotropic* speed functions, meaning that the speed at each point is direction dependent. For example, for vessel segmentation, consider a speed function that propagates the interface faster along the vessel direction than orthogonal to it [1]. An anisotropic FM method solves the Eikonal equation when the speed anisotropy is modeled as a metric tensor, i.e., in 3D, the anisotropy has the form of an ellipsoid. Anisotropic FM has been applied in medical imaging for vessel segmentation [1], brain connectivity analysis [8], and tumor growth modeling [7].

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 436–443, 2011.

In this contribution, a novel application of anisotropic FM is investigated, namely the computation of blood flow trajectories in Phase-Contrast (PC) MRI images. PC MRI is able to measure blood flow and the acquired images can be reconstructed into vector fields in which each voxel contains a vector that describes the local blood flow velocity. The clinical interest in applying PC MRI for assessing stenoses, aneurysms and heart valves, and for surgical planning in congenital heart disease, is increasing. Traditionally, streamline methods such as Runge-Kutta's are used for tracing blood flow trajectories in the PC MRI data [3], but these methods ignore the intrinsic uncertainty due to noise in the measured vector field. Recently, a sequential Monte Carlo method was proposed for investigating the *probability distribution* of flow trajectories [5]. In this work, an alternative approach to incorporate the measurement uncertainty is proposed in the framework of anisotropic minimal paths. Ideas from anisotropic Fast Marching applications in other fields, namely brain connectivity analysis [8] are adapted to the novel application to compute blood flow trajectories. As a result of the modeling of the problem as anisotropic minimal paths, any anisotropic Fast Marching method available can be applied to solve the problem.

2 Methods and Materials

2.1 Minimal Paths

Informally, a minimal path is the path along which one travels between two given points in the shortest time given a spatially varying speed function. Formally, let $\gamma(s) : \mathbb{R}_0^+ \to \mathbb{R}^3$ be a path connecting two points in 3D space, s be a positive running variable. Furthermore, let $\mathbf{M}(\mathbf{x})$ denote a metric tensor at position $\mathbf{x}$ in 3D space represented by a symmetric positive definite matrix that defines a spatially varying traveling cost, i.e., the inverse of the speed. Hence, the speed along γ is dependent both on the spatial position $\gamma(s)$ as well as the tangent direction $\gamma'(s)$. The local cost of traveling at a certain position $\gamma(s)$ along the path can be written

$$F(\gamma(s), \gamma'(s)) = \sqrt{\gamma'(s)^T \mathbf{M}(\gamma(s))\gamma'(s)} \tag{1}$$

and the total cost of a path connecting two points $\mathbf{a}$ and $\mathbf{b}$ is

$$J(\gamma) = \int_{s_\mathbf{a}=0}^{s_\mathbf{b}} F(\gamma(s), \gamma'(s))\,\mathrm{d}s = \int_{s_\mathbf{a}=0}^{s_\mathbf{b}} \sqrt{\gamma'(s)^T \mathbf{M}(\gamma(s))\gamma'(s)}\,\mathrm{d}s. \tag{2}$$

Considering all paths $\Gamma_{\mathbf{a},\mathbf{b}}$ between $\mathbf{a}$ and $\mathbf{b}$, the minimal path is the one with the minimum cost. Minimal paths from all points $\forall \mathbf{x} \in \Omega \subseteq \mathbb{R}^3$ to a target region $\partial\Omega$ can be described using the so-called value function $U(\mathbf{x})$

$$U(\mathbf{x}) = \begin{cases} \min\limits_{\gamma \in \Gamma_{\mathbf{x},\partial\Omega}} J(\gamma) & \mathbf{x} \in \Omega \\ 0 & \mathbf{x} \in \partial\Omega. \end{cases} \tag{3}$$

Along a minimal path, the values of U are strictly decreasing as we move closer in time to the target region, i.e., for a point $\mathbf{g} = \gamma(s_g)$ on the minimal path between the target region and a point $\mathbf{x} = \gamma(s_\mathbf{x})$, it holds that $U(\mathbf{g}) < U(\mathbf{x})$. More precisely,

$$U(\mathbf{x}) = U(\mathbf{g}) + \int_{s_\mathbf{g}}^{s_\mathbf{x}} F(\gamma(s), \gamma'(s))\,\mathrm{d}s, \tag{4}$$

and forms the basis for *Dijkstra's shortest path* and FM algorithms. Furthermore, it can be shown that the value function U satisfies the static Hamilton-Jacobi equation [6]

$$\nabla U(\mathbf{x})^T \mathbf{M}(\mathbf{x})^{-1} \nabla U(\mathbf{x}) = 1, \tag{5}$$

also known as the anisotropic Eikonal equation, and that the tangents of the minimal paths satisfy $\gamma' \propto \mathbf{M}^{-1}\nabla U$. Thus, we are able to solve the problem of computing the minimal paths by first solving the partial differential equation in Eq. 5 with boundary condition $U(\mathbf{x}) = 0$ on the target region $\partial\Omega$ to find $U(\mathbf{x})$, and then reconstructing the minimal paths by backtracking using standard numerical methods like Euler's, Heun's or Runge-Kutta's. It can further be noted that for an isotropic metric $\mathbf{M}(\mathbf{x}) = f(\mathbf{x})\mathbf{I}$, Eq. 5 reduces to the well known Eikonal equation $\|\nabla U(\mathbf{x})\| = \frac{1}{\sqrt{f(\mathbf{x})}}$. The numerical scheme to compute the solution of Eq. 5 with given boundary conditions $U(\mathbf{x}) = 0$ is a novel anisotropic Fast Marching solver that we developed. The description of the solver is beyond the scope of this contribution and will be done in a future publication. Generally, any anisotropic Fast Marching solver available is applicable to solve the stated problem.

2.2 Modeling Blood Flow Trajectories as Minimal Paths

Anisotropic FM has previously been applied for computing brain connectivity maps in Diffusion Tensor Imaging [8]. In this work, blood flow computation and flow connectivity mapping in PC MRI images using anisotropic FM is proposed. From the PC MRI images, a 3D vector field of blood flow velocities can be reconstructed, see Fig. 2(a). Traditionally, vector field visualization methods such as streamlines computed using Euler, Heun or Runge-Kutta schemes are used to compute blood flow trajectories [3]. However, these methods do not consider the uncertainty in the measured flow vectors due to image noise and the streamlines may therefore give a false impression of precision. In [5], it was shown that each component $\tilde{v}$ of a measured flow vector is well approximated by a Gaussian distribution around a true velocity v:

$$\tilde{v} \in N(v, \sigma) \;\text{ with }\; \sigma = \frac{v_{enc}}{\pi}\frac{\sqrt{2}}{\mathrm{SNR}}, \tag{6}$$

where v_{enc} is a known velocity encoding sequence parameter and SNR the signal-to-noise-ratio of the PC MRI images within vessels. Typical values for the parameters are $v_{enc} = 1500$ mm/s and an SNR of 10, resulting in a standard deviation of about $\sigma = 70$ mm/s. As a reference, normal flow velocities range

between 0 mm/s and 1500 mm/s at peak systole in the aorta, and in smaller vessels, the peak flow rate is lower. Each PC MRI flow vector consists of an v_x, v_y and v_z component, $\mathbf{v} = [v_x, v_y, v_z]^T$, so that a flow vector can be considered as drawn from a multivariate Gaussian distribution with independent components identically distributed according to the variance parameter σ.

To capture uncertainty in flow trajectories traced in PC MRI data, instead of considering deterministic vectors, in this work we view the data as a field of 3-dimensional multivariate Gaussian distributions. We incorporate the uncertainty of the measured velocities into a metric tensor and then solve the minimal path problem in the ensuing metric space to find flow trajectories. The inverse of the metric tensor is constructed as

$$\mathbf{M}^{-1} = \mathbf{v}\mathbf{v}^T + \sigma\mathbf{I}, \tag{7}$$

see Fig. 1 for an schematic illustration and Fig. 2(b) for a tensor field constructed based on a PC MRI data set of an aorta. Figure 2(c) shows the value function

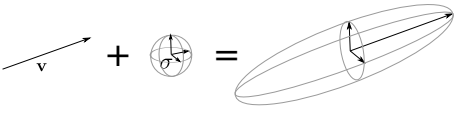


Fig. 1. Construction of a metric tensor from a given PC-MRI velocity vector

U obtained with anisotropic FM from a start point in the aortic arch. From every point in the aorta, a minimal path that represents the flow trajectory to that point can be found via a backtracking on U. To differentiate more likely paths from less likely paths, an approach similar to the one proposed in [8] in the context of DTI is adopted, in which the alignment of the minimal path tangent with the major eigenvector of the metric tensor $\mathbf{M}$ is evaluated. In the current case, the alignment can be measured by $\mathcal{C}\left(\gamma(s), \gamma'(s)\right) = \left\|\mathbf{v}\left(\gamma\left(s\right)\right)^T \gamma'\left(s\right)\right\|$ and the mean alignment $\mu\left(\mathbf{x}\right)$ along the minimal path to the point $\mathbf{x}$ is

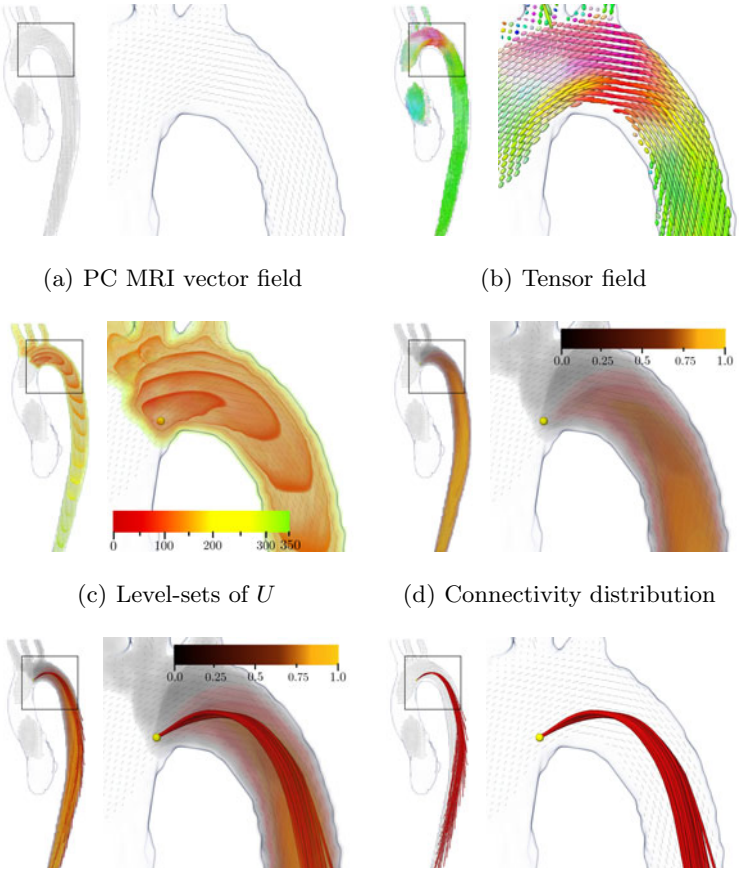
$$\mu\left(\mathbf{x}\right) = \frac{1}{U\left(\mathbf{x}\right)} \int_0^{s_{\mathbf{x}}=U(\mathbf{x})} \mathcal{C}\left(\gamma(s), \gamma'(s)\right) \mathrm{d}s. \tag{8}$$

Figure 2(d) shows $\mu\left(\mathbf{x}\right)$, which can be interpreted as a flow connectivity map. Figure 2(e) and 2(f) show the minimal paths with the highest connectivities, representing the most likely paths under the influence of noise in the PC MRI images.

Note that the PC MRI flow vectors have a direction whereas the constructed metric tensor in Eq. 7 has not. In a flow application, it may be of interest to trace trajectories either forwards or backwards. To this end, the direction of the FM propagation needs to be restricted as follows: After computation of the minimal path tangent during the FM update at a point, the propagation is restricted to points for which the dot product with the flow direction is strictly positive or negative respectively.

2.3 Data

Three *PC MRI data* sets are used to demonstrate the blood flow computation using anisotropic FM: a data set covering the aorta of a healthy subject (P1), see Fig. 2(a), a patient with an aortic aneurysm (P2), see Fig. 4, and a patient with an aneurysm in the carotid artery (P3), see Fig. 5. The aorta data sets have a spatial resolution of about $1.7 \times 1.7 \times 3\text{mm}^3$, and the carotid artery data set has a resolution of $0.86 \times 0.86 \times 1.4\text{mm}^3$.



(a) PC MRI vector field (b) Tensor field

(c) Level-sets of U (d) Connectivity distribution

(e) Connectivity distribution with most likely flow trajectories (f) Most likely flow trajectories

Fig. 2. Flow trajectories as minimal paths: (a) PC MRI vector field inside the aorta, (b) the constructed tensor field, (c) level-sets of $U(\mathbf{x})$, (d) the connectivity distribution $\mu(\mathbf{x})$, (e) the 20% most likely trajectories superimposed on the connectivity distribution, and (f) the 20% most likely trajectories inside the flow field

3 Results

The experiments were carried out on a PC with 2.4 GHz CPU and 2 GB memory. The different steps for calculating flow connectivity maps and flow trajectories as minimal paths based on PC MRI data are shown in Fig. 2 for the normal aorta data set P1. From a start point, the value function $U(\mathbf{x})$ and connectivity map $\mu(\mathbf{x})$ as described in Sec. 2.2 are calculated. The flow connectivity map may be interpreted as a distribution of possible flow trajectories under the influence of image noise. Figure 3 shows a comparison of the FM-based connectivity map with the probabilistic connectivity map computed with the method described in [5] and for which the authors kindly provided an implementation. The flow field is generated based on a sinus curve that is superimposed in the figures. The probabilistic connectivity map counts the number of pathlines passing through

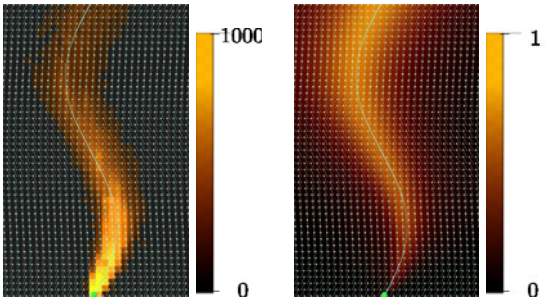


Fig. 3. Visual comparison of (a) the probabilistic connectivity map [5] and (b) the FM-based connectivity map. Superimposed in both images is the sinus curve, based on which the artificial flow field was generated, and the flow field itself. The source point is located at the bottom of the field and is marked by a sphere.

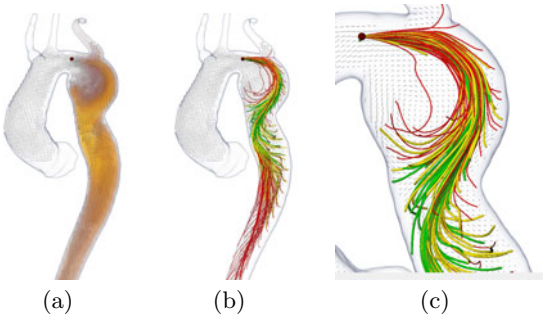
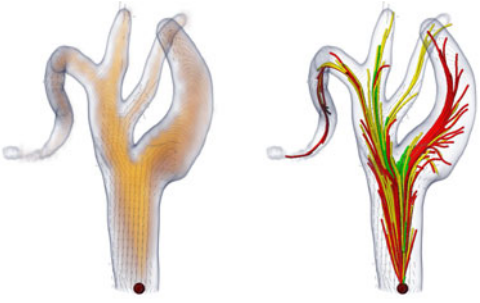


Fig. 4. Blood particle trajectories in an aorta with an aortic aneurysm: (a) Flow field and connectivity distribution, (b) blood flow trajectories, and (c) a close-up. The trajectories are color coded based on the connectivity: (green/brightest) 0.8-1.0, (yellow) 0.7-0.8, and (red/darkest) 0.6-0.7.



(a) (b)

Fig. 5. Blood flow trajectories in a carotid artery with an aneurysm: (a) flow field and connectivity distribution and (b) blood flow trajectories. Trajectories are color coded based on the connectivity: (green/brightest) 0.8-1.0, (yellow) 0.7-0.8, and (red/darkest) 0.6-0.7.

each point of the map. Both methods show a widening effect going further away from the source point that reflects uncertainty due to image noise. The major difference between both results is that the FM-based connectivity map gives the mean value of connectivity along a path, thus resulting in a almost constant value along the path, while the probabilistic connectivity map has decreasing values along a path. Figure 4 shows a visualization of the trajectories computed for data set P2 with an aortic aneurysm. The trajectories are color coded according to how likely they are: green (brightest) means a connectivity between 0.8 and 1.0, yellow between 0.7 and 0.8, and red (darkest) between 0.6 to 0.7. Similar flow paths for the carotid artery data set P3 are shown in Fig. 5. The computations for the above examples took about 10 seconds using our developed FM method.

4 Discussion and Future Work

In this contribution, a novel modeling of blood flow trajectories as minimal paths is presented. Knowing the measurement uncertainty is important in all science and decision making, including medical imaging. A visualization of blood flow as minimal paths and flow connectivity maps gives a notion of the inherent uncertainty in the PC MRI data which is not provided by traditional streamlining techniques. The focus in this contribution is to present the methodological basis for blood flow estimation based on minimal paths. Currently, the main application is uncertainty visualization for flow patterns, e.g., for surgical planning in congenital heart disease. There are several possible extensions for future work, the most immediate one being to implement the method for 4D data as the PC MRI images frequently also have a temporal dimension. Furthermore, more experimental work, using for example physical flow phantoms or flow simulations, to further validate the method is required. However, as the flow connectivity maps obtained with the anisotropic FM method do not correspond to a physical

quantity, a strict validation may prove difficult. Similar problems are encounterd in other applications, such as in the reconstruction of white matter fibers based on Diffusion Tensor Imaging, where both stochastic tracking and anisotropic FM approaches are utilized, and in which validation of the corresponding connectivity maps is still an open issue. A different avenue of work is to investigate if the flow connectivity maps can be used as an imaging biomarker for disease, for example, to investigate via group studies if connectivity-based biomarkers have any discriminative power to separate patients from healthy controls. Example applications for this include stroke and embolization pathways, flow patterns related to aneurysm rupture, and connectivities through leaking heart valves.

Acknowledgements. The authors would like to thank Dr. Michael Markl, member of Department of Radiology, Medical Physics at Universitätsklinikum Freiburg for providing the PC MRI data.

References

1. Benmansour, F., Cohen, L.D.: Tubular structure segmentation based on minimal path method and anisotropic enhancement. Int J. Comput. Vision, 1–19 (2010)
2. Cohen, L.D., Kimmel, R.: Global minimum for active contour models: A minimal path approach. Int. J. Comput. Vision 24(1), 57–78 (1997)
3. Darmofal, D., Haimes, R.: An analysis of 3D particle path integration algorithms. J. Comput. Physics 123(1), 182–195 (1996)
4. Deschamps, T., Cohen., L.D.: Fast extraction of minimal paths in 3D images and application to virtual endoscopy. Med. Image Anal. 5(4), 281–299 (2001)
5. Friman, O., Hennemuth, A., Harloff, A., Bock, J., Markl, M., Peitgen, H.O.: Probabilistic 4D blood flow mapping. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6363, pp. 416–423. Springer, Heidelberg (2010)
6. Jbabdi, S., Bellec, P., Toro, R., Daunizeau, J., Pélégrini-Issac, M., Benali, H.: Accurate anisotropic Fast Marching for diffusion-based geodesic tractography. Int. J. Biomed Imaging (2008)
7. Konukoglu, E., Sermesant, M., Clatz, O., Peyrat, J.-M., Delingette, H., Ayache, N.: A recursive anisotropic fast marching approach to reaction diffusion equation: Application to tumor growth modeling. In: Karssemeijer, N., Lelieveldt, B. (eds.) IPMI 2007. LNCS, vol. 4584, pp. 687–699. Springer, Heidelberg (2007)
8. Prados, E., Lenglet, C., Pons, J.P., Wotawa, N., Deriche, R., Faugeras, O., Soatto, S.: Control theory and Fast Marching techniques for brain connectivity mapping. In: Proceedings of CVPR 2006, pp. 1076–1083 (2006)
9. Sethian, J.A.: Level Set Methods and Fast Marching Methods Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision, and Materials Science. Cambridge University Press, Cambridge (1999)

Finding the Optimal Compression Level for Strain-Encoded (SENC) Breast MRI; Simulations and Phantom Experiments

Ahmed A. Harouni[1], Michael A. Jacobs[2], and Nael F. Osman[1,2]

[1] Electrical and Computer Engineering, Johns Hopkins University
harouni@jhu.edu
[2] Russell H. Morgan Department of Radiology and Oncology, Johns Hopkins University school of medicine, Baltimore, MD, USA

Abstract. Breast cancer is the most common cancer among women and the second highest cause of cancer-related death. Diagnostic magnetic resonance imaging (MRI) is recommended to screen high-risk patients. Strain-Encoded (SENC) can improve MRI's specificity by detecting and differentiating masses according to their stiffness. Previous phantom and ex-vivo studies have utilized SENC to detect cancerous masses. However, SENC required a 30% compression of the tissue, which may not be feasible for in-vivo imaging. In this work, we use finite element method simulations and phantom experiments to determine the minimum compression required to detect and classify masses. Results show that SENC is capable of detecting stiff masses at compression level of 7%, though higher compression is needed in order to differentiate between normal tissue and benign or malignant masses. With on-line SENC calculations implemented on the scanner console, we propose to start with small compressions for maximum patient comfort, then progress to larger compressions if any masses are detected.

1 Introduction

According to the American Cancer Society's 2009 report [1], one in eight women will develop breast cancer in her lifetime. Early detection of breast lesions using mammography has resulted in lower mortality rates. However, some breast lesions are mammography occult (e.g. dense breasts) and the use of magnetic resonance imaging (MRI) is recommended, especially for women who are at high risk of developing breast cancer.

MRI has high sensitivity (95%) and moderate specificity (83%) [2]. Cancerous masses are 3-13 times stiffer than normal tissue [3]; therefore, MRI's Specificity can be increased by incorporating the tissue's stiffness. Osman *et al.* developed Strain-Encoded (SENC) MRI [4] to directly measures strain, which is inversely proportional to stiffness. The feasibility of SENC to detect stiff masses in a homogeneous phantom was shown in [5,6] using a simple hardware with single compression that limited the scan time to one second and the scanning resolution to 4x4x10 mm^3. In [7], we introduced a new hardware capable of accurately repeating compressions, which allows to increase scanning time. Thus, achieving a high resolution of 1x1x5 mm^3 and significantly improving signal-to-noise ratio and contrast-to-noise ratio. Then in a later

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 444–451, 2011.

study [8], an alternation of SENC called SENC relaxation (SENC-REX) was introduced to complement traditional SENC compression (SENC-CMP). Phantom and ex-vivo results showed the potential of SENC-CMP and SENC-REX to detect cancerous masses. Currently however, SENC requires multiple compressions of 30%. This compression level was originally developed and optimized to image the myocardial circumferential strain, which ranges from $+5\%$ to -30% [4], but such a high compression may prove to be uncomfortable or infeasible in a clinical setting.

In this work, we focus on determining the minimum amount of compression necessary to detect and differentiate between benign and malignant masses. We built a phantom that contains masses with known stiffness, performed finite element method simulations and compared the results with acquired MRI data.

2 Methods and Material

2.1 Hardware

We used our hardware previously described in [7] for compression. The hardware consists of two air-cylinders fitted underneath a standard MR breast coil. Figure 1a shows a patient laying in prone position with two air-cylinders compressing each breast. Figure 1b shows a top-view of a single side of the hardware compressing a breast phantom. The hardware can operate in either compression or relaxation mode, generating SENC-CMP and SENC-REX strain images, which examine the compression and relaxation properties of the tissue, respectively. More details can be found in [7,8].

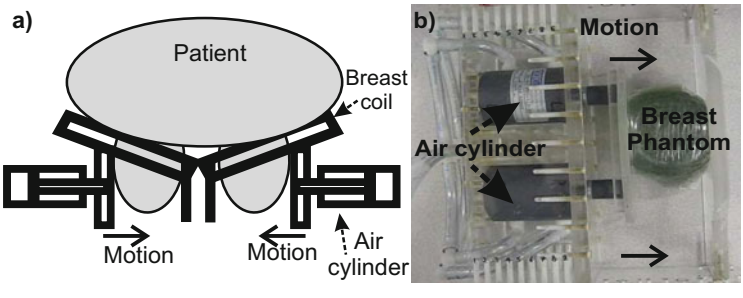


Fig. 1. a) Schematic of the hardware fitted under a standard breast coil with a patient lying in the prone position. b) Image of the hardware compressing a gel phantom.

2.2 Strain-Encoded (SENC)

As introduced in [4], SENC is a method for directly measuring strain, which is defined as the percentage change in length of tissue given by $\varepsilon = \frac{\triangle L}{L_0}$, where L_0 is the initial tissue length and $\triangle L$ is the change in length due to tissue deformation. SENC imaging applies a tagging preparation pulse, which corresponds to a sinc peak at the tagging frequency, ω_0. After deformation this peak shifts to different frequency. This shift, which is proportional to the strain, can be estimated by acquiring two images at two different z-encoding frequencies, ω_L and ω_H.

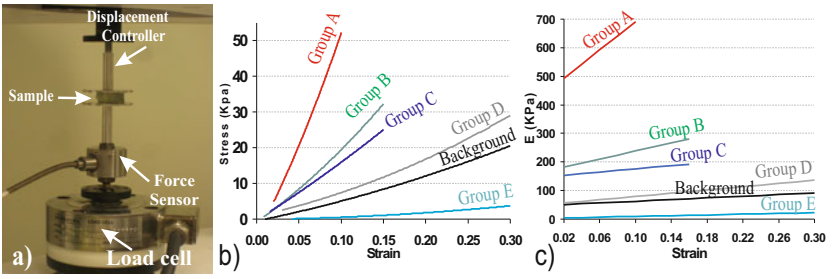


Fig. 2. a) Dynamic mechanical analyzer used to test each mass. b) Stress-strain curves for masses in Groups A, B, C, D, E and the background. c) Young's modulus (E) in KPa variation with strain for Groups A, B, C, D , E and the background.

2.3 Phantom Composition

In order to emulate breast cancer masses with different stiffnesses, a custom-made phantom was created by varying the A to B gel compounds mixing ratios of four different silicon materials. We constructed five groups of masses; masses of Groups A, B, C, and D were stiffer than the background, whereas masses of Group E were softer than the background. Table 1 shows the types of gel with corresponding mixing ratios. All of the masses were 10 mm-thick cuboids, varying in plane from 3 mm to 12 mm (Fig. 4a, b).

We used a dynamic mechanical analyzer to determine the stiffness of the phantom (see Fig. 2a). We used a ramped displacement to induce 10%-30% strain on the sample. Figure 2b shows non-linear stress-strain curves obtained from the dynamic mechanical analyzer, and Figure 2c shows Young's modulus in KPa. Groups A, B, and C are respectively 8, 3, and 2 times stiffer than the background, while the stiffness of Groups D and E are very similar to the background. Comparing our phantom's Young's modulus with the ex-vivo breast stiffness measured by Samani *et al.* [3], masses of Group A would mimic malignant tumors, Groups B and C would mimic benign masses, while Groups D and E would mimic normal tissue. Table 1 shows Young's modulus for each gel mixtures calculated as: 1) the slope of a linear fit of the stress-strain data; 2) the differentiation of the second degree polynomial fit of the stress-strain data.

Table 1. Properties of different silicon masses, including mixing ratio, corresponding Young's modulus (E) in KPa, and mass classification.

Group	Material	Mixing ratio A:B	E (KPa)		Group:Back stiffness ratio	Classification
			Linear fit	Polynomial fit		
A	3-4207	1.0:1.0	593	$E = 2478\varepsilon + 443$	8.4:1	Malignant
B	3-4133	1.5:1.0	226	$E = 714\varepsilon + 167$	3.2:1	Benign
C	3-4207	1.0:1.2	171	$E = 278\varepsilon + 148$	2.4:1	Benign
D	3-4133	1.2:1.0	100	$E = 284\varepsilon + 51$	1.4:1	Normal
Background	A-341	1.0:10.0	71	$E = 145\varepsilon + 48.5$	-	Normal
E	3-4222	1.0:1.5	20	$E = 64\varepsilon + 2.9$	0.3:1	Normal

3 Experiments

3.1 Finite Element Method (FEM) Simulations

To simulate the tissue deformation due to different compressions levels, we developed a finite element method using Matlab7.5®. We simulated a cross section of the phantom using 31x55 nodes connected together by two constant strain triangles having linear shape function. Fixed boundary conditions and Poisson ratio of 0.49 were assumed. For each compression level, we simulated a cross-section of the background containing five masses with 10 mm thick and width of 2, 4, 6, 8, and 10 mm (see Fig.3a). Multiple FEM iterations were performed to simulate the phantom's Young's modulus non-linearity.

3.2 Scanning Protocol

We performed scans on a 3T MRI Philips scanner (Achieva, Philips Medical Systems, Best, the Netherlands) using a four-channel phased array breast coil. All slices were scanned using in-plane resolution of 1x1 mm^2, field-of-view = 192 x 192 mm^2, slice thickness = 5 mm. The phantom was scanned using: T1-weighted (TR = 495 / TE = 10 ms, sense factor = 2), T2-weighted spin echo (TR = 2500/ TE = 60 ms, sense factor = 2), SENC-CMP and SENC-REX protocols. For SENC scans we used ramp flip angles described in [7] with last flip angle = 80°, Tagging delay = 100 ms, trigger delay = 500 ms, segmented Cartesian K-space acquisition using turbo field echo (TFE) factor of 10 with no echo planar imaging (EPI), 19 compression cycles/slice.

For a given expected strain value at the boundaries of the compressing device, we used SENC equations given in [9] to calculate ω_0, ω_L, and ω_H. In [9], Basha *et al.* showed that if the tissue of interest is compressed or stretched beyond the expected range, SENC measurements would saturate at either high or low strain values. Simulations showed that parts of the tissue would retain larger strain values than the strain applied at the tissue's boundaries; therefore, the scanning parameters were set to encompass strain values larger than levels applied at the boundaries (see Table 2).

Table 2. Scanning parameters for measured strain values for SENC-CMP and SENC-REX scans at different levels of applied compression. All frequencies ($\omega_0, \omega_L, \omega_H$) are in mm^{-1}.

SENC Compression (SENC-CMP)				SENC Relaxation (SENC-REX)			
Compression at boundary	Measured strain range	$\omega_0 = \omega_L$	ω_H	Relaxation at boundary	Measured strain range	ω_L	$\omega_0 = \omega_H$
-7 %	0 to -10%	1.8	2.0	+8%	0 to +10%	2.0	2.2
-10%	0 to -15%	1.132	1.332	+11%	0 to +18%	1.1111	1.3111
-16%	0 to -22%	0.708	0.908	+18%	0 to +25%	0.8	1.0
-23%	0 to -30%	0.4667	0.6667	+29%	0 to +35%	0.5714	0.7714

3.3 Quantification

Masses were manually segmented and their detectability were quantified using the CNR Elastography defined by $\mathrm{CNR}_e = \frac{2(S_{mass} - S_{background})^2}{\sigma^2_{mass} + \sigma^2_{background}}$, where S and σ are the mean and

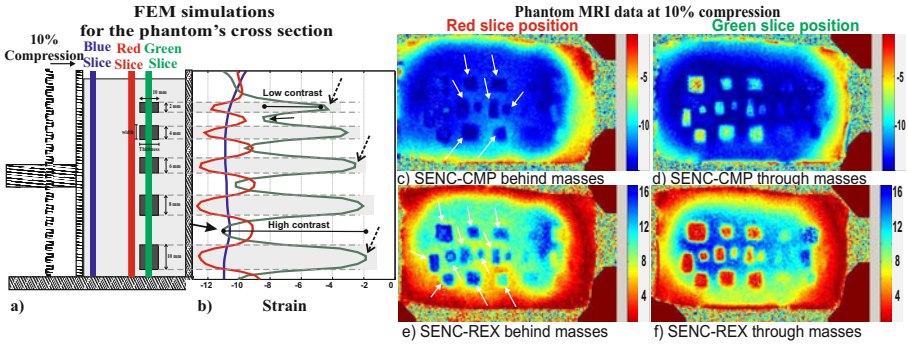


Fig. 3. a) FEM simulations while the phantom is compressed 10%. b) Strain values at blue, red and green slices calculated from the FEM simulations. Solid black arrows point to different in background strain values depending on mass separation distances. Dotted black arrows point to different strain values within the masses depending on the mass width. c-f) Phantom MRI SENC-CMP and SENC-REX images. Slice positioned immediately behind the masses (c,e) and encompassing the masses (d,f) corresponds to the red and green slices in FEM simulations. White arrows point higher strain detected behind the masses.

standard deviation, respectively, and the background is a rectangular region surrounding the mass. Philips pride software was used to generate colored SENC strain images at the scanner console immediately after acquisition. For visual comparisons,all color pallets for SENC-CMP and SENC-REX were unified such that masses having low strain values are colored red, while normal background is colored blue.

4 Results

4.1 FEM Simulation Results

Figure 3a shows a diagram for FEM simulations of a cross section of a homogeneous background with five different masses at 10% compression. All masses had same thickness (10 mm) but varied in width (2, 4, 6, 8, and 10 mm). Figure 3b shows the corresponding x-axis strain profile along blue, red, and green slices. The strain profile at the blue slice (furthest from the masses) remains constant around -11%, while the strain profile at the green slice (intersecting the masses) ranges from -2% within the masses to-11% in the surrounding background. By examining the simulation results, we observe the following:

1. The strain inside each masses depends on the width of the mass. The largest mass (10 mm in width) had strain of -2%, whereas the smallest mass (2 mm in width) had strain of -4% (see dotted black arrows in Fig. 3b).
2. The strain measured within the background between the masses depends indirectly on the distance between the masses. If the separation between masses exceeds the combined width of the two masses, then the strain approaches the expected strain (-11%). Conversely, if the separation between masses is less than the combined

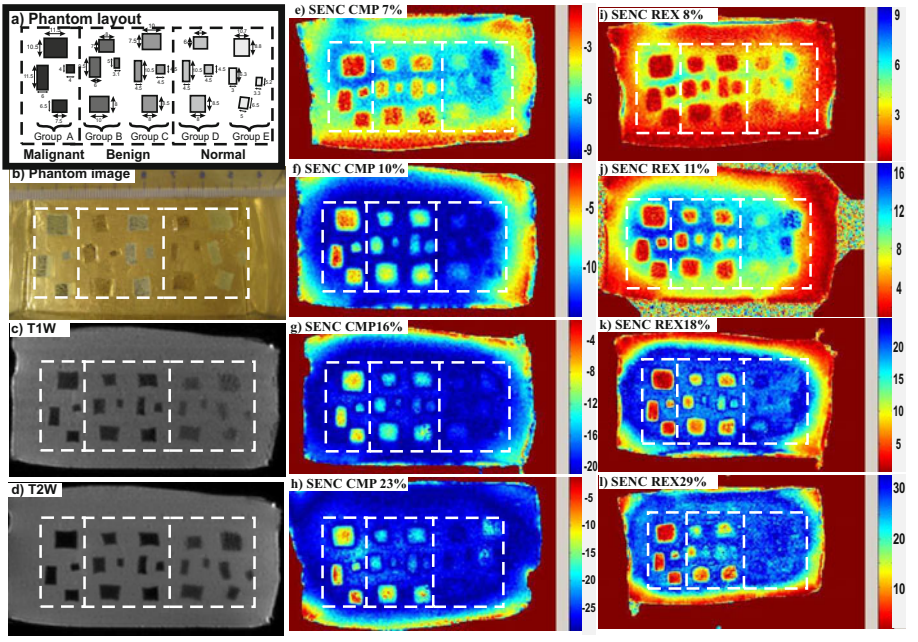


Fig. 4. a) Layout for the custom-made phantom containing five different groups of masses (dimensions are in mm). b) Image showing masses during phantom manufacturing with ruler scale. Groups A, B, C, and D are stiffer than the background, while Group E is softer than the background. c) T1W image d) T2W image for the phantom. SENC-CMP images (e-h) and SENC-REX images (i-l) for different compression levels and relaxation levels.

width of the masses, then the strain does not approach the expected strain (see solid black arrows in Fig. 3b). This interaction between masses affects the CNR_e and consequently the ability to detect the masses.

3. The strain profile of the red slice (directly adjacent to the masses) appears to be mirroring the strain profile of the green slice (encompassing the masses). This is caused by the stiff masses pushing against the surrounding soft tissue (see Fig. 3b).

4.2 Phantom Results

Figures 3c-f show SENC MRI images of the phantom at 10% compression. Figure 3d and Figure 3f are SENC-CMP and SENC-REX images, respectively, at the same position as the green slice in the FEM simulations, which encompasses the masses. Both images confirm FEM simulations results, with nearly zero strain within the masses. Figure 3c and Figure 3e show SENC-CMP and SENC-REX images, respectively, at the same position as the red slice in the FEM simulations, located adjacent to the masses. Arrows indicate the tissue surrounding the masses, which experience a higher level of compression than the applied external compression. This increases the confidence level of detecting the masses.

Figure 4c and Figure 4d show T1W and T2W images of the phantom detecting all masses: malignant (Group A), benign (Groups B and C), and normal (Groups D and E).

Table 3. Strain (mean±SD) measured from manually segmented masses form SENC-CMP and SENC-REX images at different compression and relaxation levels. Negative and positive signs indicates compression and relaxation, respectively.

	SENC Compression (SENC-CMP)				SENC Relaxation (SENC-REX)			
Compression	-7%	-10%	-16%	-23%	+8%	+11%	+18%	+29%
Group A	-3.0±1.0	-5.1±1.7	-10.5±2.6	-13.4±5.0	1.1±0.8	3.3±1.8	5.5±4.3	8.1±5.9
Group B	-3.7±0.9	-8.2±1.3	-14.5±2.0	-18.1±4.5	1.3±0.8	4.5±1.9	9.7±3.3	15.2±5.5
Group C	-3.9±0.9	-8.8±1.6	-15.0±2.4	-20.1±3.9	1.4±0.8	4.8±1.8	10.4±3.6	15.7±6.4

Figures 4e-h show SENC-CMP images for compression levels of 7%, 10%, 16%, and 23%, respectively. Figures 4i-l show SENC-REX images for relaxation levels of 8%, 11%, 18%, and 29%, respectively. Malignant masses of Group A and benign masses of Groups B and C are easily detectable visually. However, normal masses of Groups D and E lacked enough contrast to distinguish them from the background.

With SENC-CMP images at 7% compression and SENC-REX images at 8% and 11% compression, we could detect both malignant and benign but not normal masses; therefore, differentiating normal from suspicious masses. That said, these images had moderate CNR_e (11-20) and suffered from rim artifacts due to imperfect compressions. SENC-CMP images at 10% compression and SENC-REX images at 18% had higher CNR_e of 60 and 30, respectively. This allowed the differentiation between malignant and benign masses (see Table 3).

By visual comparison, relaxation always performs better than compression. None of the SENC-CMP or SENC-REX images allowed differentiation between Group B and Group C, but these groups can be distinguished from Group A for all masses larger than 4x4 mm^2 using 16% and 23% compression levels as well as 18% and 29% relaxation levels. The smallest mass (4x4 mm^2) of Group A can sometimes be misclassified as Group B and C, which matches FEM simulation results described earlier.

5 Discussion

Our phantom results demonstrate that both malignant (Group A) and benign (Groups B and C) masses can be detected under all compression and relaxation levels. With small compression (7%), SENC-CMP7% image allows the detection of malignant and benign masses with moderate CNR_e (>10). The corresponding SENC-REX8% image can also be used to detect all the masses, but has rim artifacts (see Fig. 4i). These artifacts are reduced when applying larger compression of 10% resulting in SENC-REX11% image, which has higher CNR_e and better detection capability. In order to differentiate between the benign and malignant masses we have to use even larger compression of 16% to obtain SENC-CMP16% and SENC-REX18% images.

In our phantom, the embedded masses were only eight times stiffer than the background. High-grade cancer masses are five to thirteen times stiffer than both fat and glandular tissue [3]; therefore, we would expect to see even more contrast in clinical trials. Our FEM simulations were limited to a cross-section of the phantom. If desired, this can be extended to 3D. Also, instead of using linear shape functions for the

triangular elements, which results in a constant strain for each triangular element, we could use a more complicated model.

In conclusion, we have showed using FEM simulations and phantom study that SENC-CMP and SENC-REX are able to detect both benign- and malignant-like masses with compressions as small as 7%. We have also shown that the slices adjacent to the masses show reverse contrast compared to slices containing the masses. These slices increase the confidence of our detection and classification method. Human trials need to be conducted in order to validate these phantom results. Normal volunteers, as well as patients having benign and malignant masses should be scanned. Therefore, we propose the following protocol for breast cancer screening: a) scan the whole breast using low compression. b) Examine SENC images at the scanner console using SENC on-line calculations to detect masses. c) If masses are detected, we would re-scan these slices using higher compression level in order to distinguish benign from malignant masses. Since higher compression would only be required for mass classifications, this protocol would maximize patient comfort without sacrificing detection and classification accuracy.

Acknowledgments. The authors would like to thank Shimon Unterman for his help with the stiffness measurements, as well as Issel Lim for her help in editing this paper. This work was supported in part by the following grants: NHLBI R01HL072704, NIH 1P50HL08946, 1R01CA100184, P50CA103175, P50CA88843, and U01CA140204.

References

1. American Cancer society, Cancer Facts and Figures 2009 (2009), http://www.cancer.org
2. Bluemke, D.A., Gatsonis, C., Chen, M.H., DeAngelis, G.A., DeBruhl, N., Harms, S., Heywang-Kobrunner, S.H., Hylton, N., Kuhl, C.K., Lehman, C., et al.: Magnetic resonance imaging of the breast prior to biopsy. Jama 292(22), 2735 (2004)
3. Samani, A., Zubovits, J., Plewes, D.: Elastic moduli of normal and pathological human breast tissues: an inversion-technique-based investigation of 169 samples. Physics in Medicine and Biology 52(6), 1565–1576 (2007)
4. Osman, N., Sampath, S., Atalar, E., Prince, J.: Imaging longitudinal cardiac strain on short-axis images using strain-encoded MRI. Magnetic Resonance in Medicine 46(2), 324–334 (2001)
5. Osman, N.: Detecting stiff masses using strain-encoded (SENC) imaging. Magnetic Resonance in Medicine 49(3), 606–608 (2003)
6. Fahmy, A., Krieger, A., Osman, N.: An integrated system for real-time detection of stiff masses with a single compression. IEEE Transactions on Biomedical Engineering 53(7), 1286–1293 (2006)
7. Harouni, A.A., Hossain, J., Jacobs, M.A., Osman, N.F.: Improved Hardware for Higher Spatial Resolution Strain-Encoded (SENC) MRI. Academic Radiology 18, 705–715 (2011)
8. Harouni, A.A., El Khouli, R.H., Hossain, J., Bluemke, D.A., Osman, N.F., Jacobs, M.A.: Enhancing mass detection and classification in breast tissue using Strain-ENCoded (SENC) breast MRI with histological validation. Submitted to Journal of Magnetic Resonance Imaging (March 2011)
9. Yousef, T.A., Osman, N.F.: Effect of Noise and Slice Profile on Strain Quantifications of Strain Encoding (SENC) MRI. In: Sachse, F.B., Seemann, G. (eds.) FIHM 2007. LNCS, vol. 4466, pp. 50–59. Springer, Heidelberg (2007)

Towards Patient-Specific Finite-Element Simulation of MitralClip Procedure

T. Mansi[1], I. Voigt[1,2], E. Assoumou Mengue[1], R. Ionasec[1], B. Georgescu[1], T. Noack[3], J. Seeburger[3], and D. Comaniciu[1]

[1] Siemens Corporate Research, Image Analytics and Informatics, Princeton, NJ, USA
[2] Friedrich-Alexander-University, Pattern Recognition Lab, Erlangen, Germany
[3] Department of Cardiac Surgery, Heart Centre Leipzig, Leipzig, Germany

Abstract. MitralClip is a novel minimally invasive procedure to treat mitral valve (MV) regurgitation. It consists in clipping the mitral leaflets together to close the regurgitant hole. A careful preoperative planning is necessary to select respondent patients and to determine the clipping sites. Although preliminary indications criteria are established, they lack prediction power with respect to complications and effectiveness of the therapy in specific patients. We propose an integrated framework for personalized simulation of MV function and apply it to simulate MitralClip procedure. A patient-specific dynamic model of the MV apparatus is computed automatically from 4D TEE images. A biomechanical model of the MV, constrained by the observed motion of the mitral annulus and papillary muscles, is employed to simulate valve closure and MitralClip intervention. The proposed integrated framework enables, for the first time, to quantitatively evaluate an MV finite-element model *in-vivo*, on eleven patients, and to predict the outcome of MitralClip intervention in one of these patients. The simulations are compared to ground truth and to postoperative images, resulting in promising accuracy (average point-to-mesh distance: $1.47 \pm 0.24\, mm$). Our framework may constitute a tool for MV therapy planning and patient management.

1 Introduction

The mitral valve (MV), between the left atrium and the left ventricle, prevents the blood from coming back to the left atrium during systole. Incorrect MV closure appears in many cardiac diseases and often requires surgery. The edge-to-edge technique, which consists in suturing the two mitral leaflets at the regurgitant hole, has demonstrated good clinical outcomes in patients with severe mitral insufficiency due to leaflet prolapse or calcified annulus [7]. Nowadays, this procedure can be performed percutaneously by clipping the leaflets using a MitralClip catheter [3]. Nonetheless, a careful preoperative planning is necessary to select respondent patients and to determine the clipping sites. Although preliminary indications criteria are established, they lack prediction power with respect to complications and effectiveness of the therapy in specific patients. It is not uncommon to perform several trials during the intervention and, in some

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 452–459, 2011.

cases, decide to place two clips ($\approx 30\%$ of the patients [3]) or even to abort the procedure due to complications ($\approx 10\%$ of the patients [3]). Therefore, there is a need for an efficient and predictive framework that can assist the surgeon in planning the MitralClip procedure and guide him during the intervention.

Driven by the growing prevalence of MV diseases, researchers are developing computational models of MV biomechanics to simulate its function. Several constitutive laws have been proposed, from simple isotropic linear elasticity to more complex anisotropic non-linear hyper-elasticity [9,11,13]. Fluid-structure interaction models have also been investigated [2]. Yet, most of these models have been developed on synthetic or *ex-vivo* anatomies [4]. Patient-specific anatomies and boundary conditions are starting to be used but tedious manual delineations are still required, with no [12] or partial automation [1,13]. Recently, a patient-specific simulation of MV annuloplasty has been presented [12] but the results were not confronted to postoperative data. At the same time, automatic algorithms are being developed to delineate the MV in medical images. In [5], the authors provided a fast and accurate method based on machine learning to detect the MV on 3D+t transesophageal echocardiogram (TEE) or CT images. All the elements are starting to be available for patient-specific MV simulations.

We thus propose in this paper to combine data-driven modeling of the MV with a biomechanical model of the valve apparatus to simulate MV closure in patients and test therapies, in particular the MitralClip procedure. As described in Sec. 2, we use machine-learning techniques to automatically detect the complete MV apparatus in sequences of 3D TEE images. A biomechanical model of the MV is then employed to simulate valve closure on the patient-specific anatomies and boundary conditions. As model generation is automatic and integrated, we could quantitatively evaluate the simulations with respect to the observed valve motion in eleven subjects (Sec. 3). In one of these patients, post- MitralClip images were available. The intervention was simulated and compared with the real outcome, showing promising prediction power.

2 Methods

Starting from 4D TEE images, we automatically detect the MV apparatus in all time frames using machine-learning algorithms and generate a patient-specific anatomical model of the open MV (Sec. 2.1). We then apply a biomechanical model to simulate valve closure (Sec. 2.2) and valve clipping (Sec. 2.3).

2.1 Anatomical Model of the Mitral Valve Apparatus

The anatomical model of the MV apparatus that is detected in the images comprises (Fig. 1): the mitral annulus, the anterior and posterior leaflets (AL and PL resp.), the anterior and posterior papillary heads and the chordae. To capture a broad spectrum of morphological variations, the model is parameterized by three coarse-to-fine components: *i)* Three transforms B for global location, orientation and scale over the cardiac cycle; *ii)* The trajectories of ten anatomical landmarks

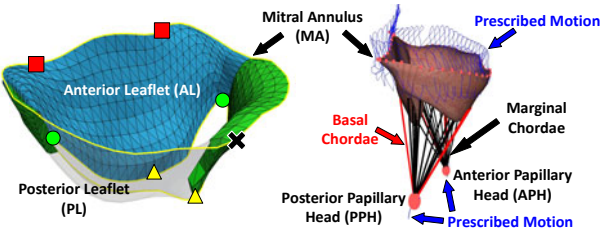


Fig. 1. Anatomical model of the MV and subvalvular apparatus, with 2 trigones *(squares)*, 1 posterior annulus mid-point *(cross)*, 2 commissures *(circles)* and 2 leaflet tips *(triangles)*. The tetrahedral mesh (left) is tagged for regional tissue properties.

$L(B) = (\mathbf{l}_1 \dots \mathbf{l}_{10}) \in \mathbb{R}^{3\times 10}$ (see Fig. 1); and *iii)* Two triangulated surface meshes $S_{LA}(B, L)$ and $S_{LV}(B, L)$ to represent the left atrial (LA) and left ventricular (LV) leaflet surfaces respectively. The vertices of each surface are constrained by the landmarks, resulting in an anatomically consistent parameterization that ensures intra- and inter-patient point correspondence.

B, $L(B)$ and $S_{LA}(L, B)$ are personalized from the images using a hierarchical discriminative learning algorithm [5]. The probability $p(B, L, S|I)$ knowing the image data I is incrementally modeled within the Marginal Space Learning (MSL) framework, based on the Probabilistic Boosting Tree (PBT) [14]. Due to the poor image quality, it is still difficult to estimate the thickness of the leaflets automatically and reliably at every time frame of the cardiac sequence. We therefore obtain the LV leaflet surface S_{LV} by artificially extruding S_{LA} in the direction of the surface normals, towards the LV, by $2\,mm$, the average leaflet thickness measured in our patients (see Sec. 3).

Finally, we generate a tetrahedral volume mesh of the MV (Fig. 1). For regional personalization, the tetrahedra are tagged according to the leaflet they belong to thanks to the anatomically consistent surface parameterization. 30 marginal chordae are evenly attached between papillary heads and leaflet free edges as well as four sets of two basal chordae, two sets for each leaflet. Insertion points, identical for every patient according to the point correspondence inherited from the anatomical model, are determined by visual inspection of the images, when visible, or like in previous studies [4,13] otherwise.

2.2 Biomechanical Model of Mitral Valve Apparatus

Valve closure is simulated by solving the dynamic system $\mathrm{M}\ddot{\mathbf{U}} + \mathrm{C}\dot{\mathbf{U}} + \mathrm{K}\mathbf{U} = \mathbf{F}_c + \mathbf{F}_p$. $\mathbf{U}$ is the displacement vector of the free vertices of the MV mesh, $\dot{\mathbf{U}}$ their velocity and $\ddot{\mathbf{U}}$ their acceleration. M is the diagonal mass matrix (leaflet mass density $\rho = 1.04\,g/mL$), K is the stiffness matrix of the internal elastic forces and C is a Rayleigh damping matrix with coefficient 0.1 for both M and K. $\mathbf{F}_c$ and $\mathbf{F}_p$ are the forces developed by the chordae and heart pressure respectively.

Leaflets are near-incompressible, anisotropic, non-linear elastic [10]. In this study we are not directly interested in leaflet stresses but we rather seek to

predict how well they close to assess possible residual regurgitant holes after MitralClip intervention. Hence, we approximate the MV properties by a linear isotropic elastic model [4], which optimizes the computational efficiency for fast simulations in the operative room. Near-incompressibility is achieved with a Poisson ratio ν of 0.488. The AL being stiffer than the PL, two different Young moduli are used, $\mathrm{E}_{AL} = 2.08\,MPa$ and $\mathrm{E}_{PL} = 1.88\,MPa$ respectively [11]. In this study we considered the cross-fiber stiffness (perpendicular to the mitral annulus), lower than the fiber stiffness, to capture the radial deformations.

Chordae are modeled by piecewise tensile springs between papillary heads and insertion points (Fig. 1), $\mathbf{F}_{c,i} = -k_{c,i}(\epsilon_{c,i}) \times (L_i - L_{i,0})$, $i \in \{marginal, basal\}$. L_i is the current elongation and $L_{i,0}$ is the rest length defined as the distance between the papillary heads and the insertion points measured at mid-diastole. The stiffness $k_{c,i}(\epsilon_{c,i})$ depends on the strain $\epsilon_{c,i} = (L_i - L_{i,0})/L_{i,0}$ to model the non-linear response of the chordae. At compression, $\epsilon < 0$, $k_{c,i} = 0\,g/mm$ (free compression). At low tension $\epsilon < 2.5\%$, chordae exhibit low stress-strain behavior (Young moduli $E_{c,basal} = 66\,g/mm^2$, $E_{c,marginal} = 312\,g/mm^2$), which then increases dramatically and almost linearly ($E_{c,basal} = 2120\,g/mm^2$, $E_{c,marginal} = 3406\,g/mm^2$) [6]. Spring stiffnesses $k_{c,i}$ are calculated from the chordae Young moduli by $k_{c,i} = A_{0,i}E_{c,i}/L_{0,i}$. $A_{0,i}$ is the chordae cross-section at rest ($A_{0,basal} = 2.05\,mm^2$, $A_{0,marginal} = 0.40\,mm^2$).

Pressures being not available, we apply a generic profile that increases from $0\,mmHg$ to $120\,mmHg$ [9]. The motion of the papillary heads, modeled as spatial points, and of the mitral annulus is prescribed from the automatic detection. This contribution is of fundamental importance as valve closure highly depends on the papillary positions and the shape of the annulus during systole [9]. As mitral clip modifies the leaflets morphology only, it is reasonable to assume that the acute motion of the annulus and papillaries stays unchanged (it depends mostly on the ventricular mechanics). Self-collisions and frictions are handled using collision springs.

The models are implemented in SOFA[1], a real-time soft-tissue intervention platform. The dynamic system is solved using co-rotational tetrahedral finite elements to cope with large deformations and rotations [8]. The simulation time is personalized and scaled such that the simulated MV closure is $10\times$ longer than what is observed in the images (from $70\,ms$ to $150\,ms$) to handle the strong and discontinuous contact forces. An implicit Euler solver is used to update mesh positions. We finally stress that the pipeline is completely integrated and automatic, although the user can manually adjust the models if necessary.

2.3 MitralClip Simulation

Virtual mitral clipping is performed interactively on the preoperative anatomical model as illustrated in Fig. 2. Stiff springs ($k_{clip} = 1000\,g/mm$) are created between the two leaflets to simulate the clip. Tissue properties are like in Sec. 2.2.

[1] http://www.sofa-framework.org

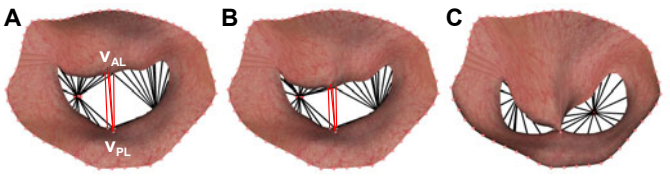


Fig. 2. Virtual MitralClip procedure. (**A**) The user picks two vertices $\mathbf{v_{AL}}$ and $\mathbf{v_{PL}}$ on each leaflet to create springs (*red lines*) that will (**B**) progressively bring the leaflets close to each other and (**C**) stitch them. The procedure is done interactively.

3 Experiments and Results

Validation of the MV Anatomical Model. Automatic detection of the MV and its subvalvular apparatus was quantitatively validated on 200 4D TEE images from 120 patients with various diseases (MV prolapse, calcified mitral annulus, stenosis, ventricular dysfunction, ...). Images were acquired with different capture range and image resolutions. Three-fold cross-validation against manual delineation yielded a point-to-mesh error of $2.75 \pm 0.86\,mm$ (detection speed: $4.8\,s$ per 3D volume, Intel Core2Duo, 2.66GHz quad core, 2GB RAM). As shown in Fig. 3, the model was able to faithfully track the MV even during valve closure.

Evaluation of MV Closure Simulation. Our simulation framework was evaluated on eleven randomly selected patients from three hospitals with various degrees of MV regurgitation or MV stenosis. The complete MV apparatus was detected, under expert guidance, on clinical 4D TEE images (image resolution: $0.75-1.58\,mm$ isotropic, $8-23$ time frames). Papillary heads could be detected reliably as they were visible on all images. The anatomical model at end-diastole, when the valve is fully open, was used to simulate MV closure. Tetrahedral meshes showed no skewed cells. The detected motion of the mitral annulus and papillary heads were used as boundary conditions. Because the biomechanical parameters of the leaflets and chordae are difficult to identify in clinical data, we decided to use the nominal parameters reported in Sec. 2 for every patient. This enabled us to evaluate the generalization of such an approach for future clinical use. Table 1 reports the average point-to-mesh distances from the simulations to the

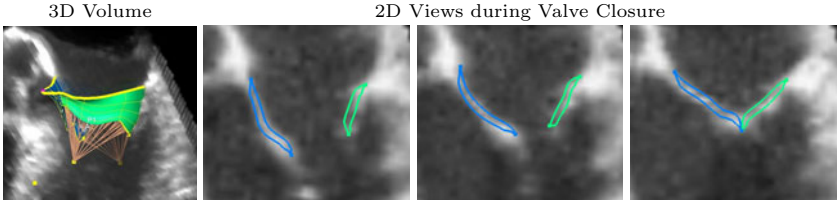


Fig. 3. Automatic detection of the mitral valve on a time sequence of 3D TEE. As one can see, the detected model faithfully tracks the moving valve over time.

Table 1. Point-to-mesh distance between simulated and detected closed configuration of the MV valve. In average, the error ($1.47 \pm 0.24\,mm$) was of the same order of magnitude as that of the automatic detection ($2.75 \pm 0.86\,mm$).

Patient	Point-to-Mesh Error	Patient	Point-to-Mesh Error	Patient	Point-to-Mesh Error
01	$1.33 \pm 0.77\,mm$	05	$1.56 \pm 1.24\,mm$	09	$2.00 \pm 1.62\,mm$
02	$1.16 \pm 0.68\,mm$	06	$1.36 \pm 0.82\,mm$	10	$1.53 \pm 1.42\,mm$
03	$1.56 \pm 1.14\,mm$	07	$1.36 \pm 0.89\,mm$	11	$1.25 \pm 0.80\,mm$
04	$1.31 \pm 0.81\,mm$	08	$1.76 \pm 1.40\,mm$		

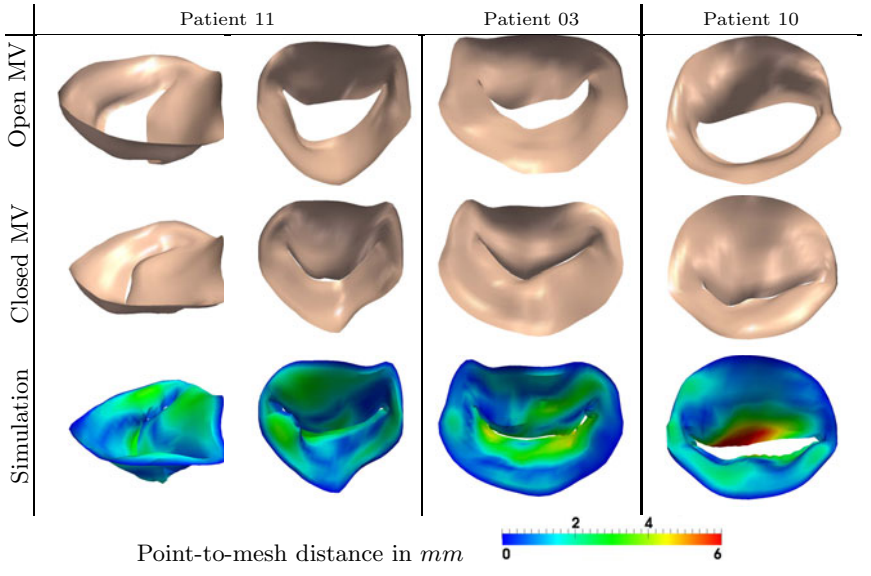


Fig. 4. Simulated MV closure in three patients compared to observed MV shape. In seven out of eleven cases, the biomechanical model managed to simulate valve closure, as in patient 03 and 11. In four patients, some chordae were too short to achieve a correct closure, as in patient 10. See text for details.

anatomical models detected on the first frame when the MV is closed. The average error was $1.47 \pm 0.24\,mm$, which is of the same order of magnitude of the automated detection. The biomechanical model correctly reproduced valve closure in seven patients (63%) despite the simplified leaflet model and the generic parameters (Fig. 4, left and mid panels). On the other four patients, the valve did not coapt correctly (Fig. 4, right panel). Nonetheless, correct coaptation could be simulated by adjusting the rest length of selected chordae, which confirms their importance in MV function [9]. Simulation speed was ≈ 4 frames per second (fps). Total simulation time was $\approx 20\,s$ with a time-step $\Delta t = 10\,ms$ (Intel Core2Duo, 2.66GHz dual core, 4GB RAM). Non-reported simulations with $\Delta t = \{1\,ms, 0.1\,ms\}$ yielded very similar results, confirming the temporal convergence of the simulation. The entire process, from the TEE images to the simulation, took about one minute.

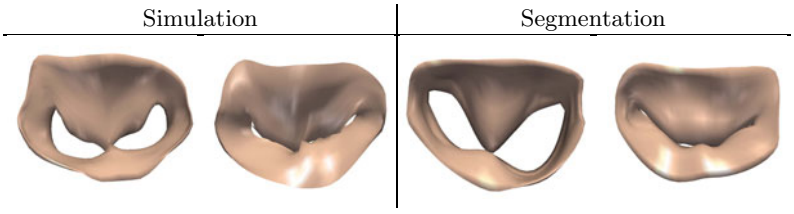


Fig. 5. To assess the ability of the model to predict MitralClip outcomes, the intervention was simulated on the preoperative anatomy of a patient (Fig. 4, patient 1), with preoperative boundary conditions (*left panel*). The result was qualitatively consistent with the true postoperative outcome (*right panel*).

Simulation of Mitral Clip. We tested the ability of our model to predict the acute outcome of MitralClip intervention in one patient (patient 3) for whom intra-operative images just after clip release were available. Since the simulation of valve closure on the preoperative anatomy was successful, we could reproduce the intervention on the preoperative anatomy and simulate the subsequent valve closure (Fig. 2, frame-rate: 8 fps) using the preoperative biomechanical parameters and boundary conditions (mitral annulus and papillary heads motion). Results were qualitatively similar to the real surgical outcome (Fig. 5).

4 Discussion and Future Works

We proposed in this paper an integrated framework for the personalized simulation of MV closure and MitralClip procedure. Our approach, fully automatic, enables easy and computationally efficient simulations. We could evaluate a patient-specific biomechanical model of the MV in eleven patients. Results demonstrated that despite the simplifications of our model and the generic parameters, good prediction power can be achieved by using patient-specific anatomies and mitral annulus and papillary heads motion. We also simulated MitralClip intervention in one patient, resulting in promising predictions compared to postoperative data. In some patients, valve closure could be achieved only by adjusting chordae biomechanical parameters (rest length and stiffness). It is therefore important to model chordae correctly for accurate and predictive simulations. As a first step, we proposed to estimate chordae rest length from the mid-diastole images, when they are at rest. One could use inverse problem methods to further personalize them, along with MV stiffness. Yet, detecting the insertion points remains a technical challenge. The proposed framework can be easily extended to other imaging modalities like TTE or CT, wherever the MV is visible. Future works include the development of non-linear anisotropic model of MV [13], simulation of valve opening to assess the effects of MitralClip on the diastolic function, anatomical model improvement and validation on larger cohorts. To the best of our knowledge, it is the first time that a biomechanical model of the MV is evaluated against *in-vivo* clinical data in a predictive manner. As illustrated by our MitralClip simulation, our framework could open, once validated, new perspectives to test MV therapies *in-silico* to optimize treatment outcome.

Acknowledgement. The authors warmly thank E. Pernod for his help with SOFA.

References

1. Burlina, P., Sprouse, C., DeMenthon, D., Jorstad, A., Juang, R., Contijoch, F., Abraham, T., Yuh, D., McVeigh, E.: Patient-specific modeling and analysis of the mitral valve using 3d-tee. In: Proc. IPCAI, pp. 135–146 (2010)
2. Einstein, D., Del Pin, F., Jiao, X., Kuprat, A., Carson, J., Kunzelman, K., Cochran, R., Guccione, J., Ratcliffe, M.: Fluid-structure interactions of the mitral valve and left heart: Comprehensive strategies, past, present and future. International Journal for Numerical Methods in Biomedical Engineering 26(3-4), 348–380 (2010)
3. Feldman, T., Kar, S., Rinaldi, M., Fail, P., Hermiller, J., Smalling, R., Whitlow, P., Gray, W., Low, R., Herrmann, H., et al.: Percutaneous mitral repair with the mitraclip system: Safety and midterm durability in the initial everest (endovascular valve edge-to-edge repair study) cohort. JACC 54(8), 686–694 (2009)
4. Hammer, P., Vasilyev, N., Perrin, D., Del Nido, P., Howe, R.: Fast image-based model of mitral valve closure for surgical planning. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 15–26. Springer, Heidelberg (2008)
5. Ionasec, R., 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 4D cardiac CT and TEE. IEEE TMI 29(9), 1636–1651 (2010)
6. Kunzelman, K., Cochran, K.: Mechanical properties of basal and marginal mitral valve chordae tendineae. ASAIO Journal 36(3), M405 (1990)
7. Maisano, F., Torracca, L., Oppizzi, M., Stefano, P., d'Addario, G., La Canna, G., Zogno, M., Alfieri, O.: The edge-to-edge technique: a simplified method to correct mitral insufficiency. European Journal of Cardio-thoracic Surgery 13(3), 240 (1998)
8. Nesme, M., Payan, Y., Faure, F.: Efficient, physically plausible finite elements. In: Eurographics (short papers), pp. 77–80 (2005)
9. Prot, V., Haaverstad, R., Skallerud, B.: Finite element analysis of the mitral apparatus: annulus shape effect and chordal force distribution. Biomechanics and Modeling in Mechanobiology 8(1), 43–55 (2009)
10. Sacks, M., Yoganathan, A.: Heart valve function: a biomechanical perspective. Phil. Trans. R. Soc. B 362(1484), 1369 (2007)
11. Schievano, S., Kunzelman, K., Nicosia, M., Cochran, R., Einstein, D., Khambadkone, S., Bonhoeffer, P.: Percutaneous mitral valve dilatation: Single balloon versus double balloon. a finite element study. J. Heart Valve Disease 18, 28–34 (2009)
12. Stevanella, M., Maffessanti, F., Conti, C., Votta, E., Arnoldi, A., Lombardi, M., Parodi, O., Caiani, E., Redaelli, A.: Mitral valve patient-specific finite element modeling from cardiac MRI: Application to an annuloplasty procedure. In: Cardiovascular Engineering and Technology, pp. 1–11 (2011)
13. Votta, E., Caiani, E., Veronesi, F., Soncini, M., Montevecchi, F., Redaelli, A.: Mitral valve finite-element modelling from ultrasound data: a pilot study for a new approach to understand mitral function and clinical scenarios. Phil. Trans. R. Soc. A 366(1879), 3411 (1879)
14. Zheng, Y., Barbu, A., Georgescu, B., Scheuering, M., Comaniciu, D.: Four-chamber heart modeling and automatic segmentation for 3D cardiac CT volumes using marginal space learning and steerable features. IEEE TMI 27, 1668–1681 (2008)

Model-Based Deformable Registration of Preoperative 3D to Intraoperative Low-Resolution 3D and 2D Sequences of MR Images

Bahram Marami, Shahin Sirouspour, and David W. Capson

Department of Electrical and Computer Engineering, McMaster University,
1280 Main Street West, Hamilton, ON, Canada, L8S-4K1
{maramib,capson}@mcmaster.ca, sirouspour@ece.mcmaster.ca

Abstract. We have developed an automatic model-based deformable registration method applicable to MR soft-tissue imaging. The registration algorithm uses a dynamic finite element (FE) continuum mechanics model of the tissue deformation to register its 3D preoperative images with intraoperative 1) 3D low-resolution or 2) 2D MR images. The registration is achieved through a filtering process that combines information from the deformation model and observation errors based on correlation ratio, mutual information or sum of square differences between images. Experimental results with a breast phantom show that the proposed method converges in few iterations in the presence of very large deformations, similar to those typically observed in breast biopsy applications.

Keywords: Image registration, soft-tissue deformation, MR imaging, breast biopsy, continuum mechanics, finite element method.

1 Introduction

Image registration is the process of aligning a pair of images to establish special correspondence between their features. Co-registering images obtained through multiple modalities or over a period of time can be instrumental in the diagnosis of diseases. Modern computed tomography (CT) and magnetic resonance imaging (MRI) systems can precisely image individual anatomy of human organs. However, plans based on preoperative images are not always effective during surgery because of possible movement and deformation of underlying tissue. In the last two decades, a considerable amount of effort has been devoted to research on developing intraoperative imaging techniques [1]. Although interventional MRI (iMRI) systems provide intraoperative images, their long acquisition time restricts the number of 2D slices that can be taken during the operation. A single or even a few slices of data provide much less information than a volume of high-resolution data. Furthermore, iMRI images often have lower signal-to-noise ratio than diagnostic MR images [1].

Finite element method (FEM) discretization of the continuum mechanics based model using elastic body deformation is the most popular physical model-based analysis in various medical applications. This method is more accurate

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 460–467, 2011.

and reliable than other simpler methods such as mass-spring modeling [2]. FEM-based deformation analysis has been extensively studied for surgical simulation especially for modeling of brain shift and deformation in neurosurgery [3]. In breast imaging, FE models have been used to predict mechanical deformations during MRI-guided biopsy [4], and to validate non-rigid registration algorithms. Moreover, they have been explored to model multi-object deformations in abdominal region for deformable image registration [5]. Similarities (or differences) between two sets of image data to be matched are used to deform the FE model utilized in non-rigid registration. In most cases, local information in images such as objects' surfaces and extracted feature points are employed to find external and internal forces applied to a discretized mesh [3]. In [4], landmarks are tapped to the breast surface in order to track its movement during biopsy. One main drawback of these methods is that they require user intervention. Another shortcoming is that surface construction via edge detection and feature extraction is a difficult and unreliable process especially in multi-modal image matching.

In this paper we propose a FE model based deformable registration method to volumetrically register high-resolution undeformed MR images of a breast phantom to low-resolution deformed data. We also provide an algorithm to non-rigidly register 3D undeformed images to 2D slices of the deformed data. In the proposed method a *dynamic* FE model acts as a regularization constraint on the image similarity criterion. The model not only leads the nonlinear optimization problem involved in registration process to its global extremum, but also provides temporal correlation for information obtained through 2D slices over time. This would allow tracking of the 3D volume deformation of soft-tissue based on intraoperative 2D sequences of MR images using a filtering process. In the volumetric registration, a voxel intensities based variational method is used to implicitly compute the forces applied to the 3D model. The same method is employed to find displacements of the regular 2D grid of the preoperative volume to match it with the 2D intraoperative image.

2 Methods

2.1 Mathematical Formulation of the Registration Problem

Given a reference image R and a template image T, image registration is basically finding a reasonable displacement field u, so that the transformed image $T[u]$ is similar to R. The objective function to be minimized is:

$$J(u) = I(T[u], R) + \alpha S(u); \quad \alpha \in \Re_+ \tag{1}$$

where I is the distance (similarity) measure between two images. In this paper, three different measures are considered: sum of squared differences (SSD) between the images which is a distance measure, mutual information (MI) and correlation ratio (CR) of two image data which are both similarity measures. In (1), S is a regularization term which ensures that the displacement field is "reasonable". The linear elastic energy of the deformable body is used as the

regularization term in our work [3,4,5]. Finally α weighs the importance of the regulation term compared with the distance (similarity) measure.

In a linear elastic continuum with no initial strains or stresses the potential energy of a body subjected to externally applied forces can be expressed as [6]:

$$E = \int_{\Omega} \sigma^T \epsilon \, d\Omega + \int_{\Omega} u^T f \, d\Omega \tag{2}$$

where f is the vector of forces applied to the elastic body, u the displacement field, and Ω is the body of the elastic subject. Also, ϵ and σ are the strain and stress vectors respectively which are explained in details in [6,7]. Based on the concept of a FE discretization, a volume of elastic body is approximated as an assemblage of discrete finite elements interconnected at nodal points on the element boundaries. Hence, displacements within the elastic body can be written as a function of the displacements at the element's nodal points weighted by the element's shape function [7]. Discretizing the continuous body with tetrahedral elements, for the volume of every tetrahedral element el with 4 vertices points, the potential energy function at every node i (2) can be written as:

$$E(u_i^{el}) = \int_{\Omega} \sum_{j=1}^{4} {u_i^{el}}^T {B_i^{el}}^T D B_j^{el} u_j^{el} \, d\Omega + \int_{\Omega} {u_i^{el}}^T f_i^{el} \, d\Omega \tag{3}$$

Here u_i^{el} is the nodal points displacement, f_i^{el} is the vector of forces concentrated at the nodal points, D is the elasticity matrix characterizing the material's property, and B_i^{el} depends on the shape function [6]. f_i^{el} can be computed as a classical optical flow field between the images to be matched; this would provide a semi-implicit method where the optical flow field would be an initial estimate of the deformation field which is regularized by the elastic model [7]. In the discretized domain, the objective function to be minimized (1) would become:

$$J(u_i^{el}) = I(T[u_i^{el}], R) + \alpha \int_{\Omega} \sum_{j=1}^{4} {u_i^{el}}^T {B_i^{el}}^T D B_j^{el} u_j^{el} \, d\Omega \tag{4}$$

If J has a local extremum at u_i^{el}, its first variation at u_i^{el} must vanish, i.e. $\frac{dJ(u_i^{el})}{du_i^{el}} = 0$. Then using Eq. (4), we have:

$$\frac{dI(T[u_i^{el}], R)}{du_i^{el}} + \alpha \int_{\Omega} \sum_{j=1}^{4} {B_i^{el}}^T D B_j^{el} u_j^{el} \, d\Omega = 0 \tag{5}$$

Comparing (3) and (4), and considering forces f_i^{el} concentrated at the nodal points, (5) can be written in a matrix form for each element as $K^{el} u^{el} = f^{el}$ where $K_{i,j}^{el} = \int_{\Omega} {B_i^{el}}^T D B_j^{el} \, d\Omega$ is a 3×3 matrix and every element i,j refers to pairs of nodes of the element el; $f_j^{el} = -\frac{1}{\alpha} \frac{dI(T[u_j^{el}], R)}{du_j^{el}}$ is a 3×1 vector. The

12×12 matrices K^{el} and the vector f_j^{el} are computed for each element and are then assembled in a global system as:

$$Ku = f(u) \tag{6}$$

where K is the global stiffness matrix associated with the volumetric mesh, u is the vector of nodal displacements and $f(u)$ is the vector of nodal forces. The solution to this nonlinear system of equations will provide the displacement field corresponding to the global minimum of the objective function (1).

2.2 Dynamic Finite Element Model

An iterative numerical method has to be employed to solve the nonlinear system of equations in (6). To this end, in this paper we consider a second-order dynamical system of the form of [8]

$$M\ddot{u} + C\dot{u} + Ku = f(u) \tag{7}$$

where M is the mass matrix of the elements concentrated at nodes, and $C = \beta M + \gamma K$ is the damping matrix for constant values of β and γ. It is noted that the steady-stated equilibrium of the this dynamical system is the solution to the static system of equations in (6). The dynamic equations can be solved using existing implicit or explicit numerical integration routines over time.

This dynamic model also allows for real-time intraoperative registration of a deforming organ in our approach. It provides a temporal correlation model for the images taken at different sample times. Such situation can arise, for example, in real-time MR based biopsy interventions where the soft tissue undergoes deformation due to the force of needle insertion. In this paper a modal-based model reduction is employed to solve (7) more efficiently [8]. Using this method, very fast modes of (7) are eliminated to significantly reduce computations without affecting the steady-state solution.

2.3 3D Volumetric Registration

In the volumetric registration, 3D high-resolution and 3D low-resolution MR images of an undeformed and deformed breast phantom are the template and reference image data sets, respectively. The nodal forces are computed based on the derivative of the distance (similarity) measure. By letting h denote a generic intensity comparison function, $f(u)$ at each iteration can be computed as:

$$f(u) = -\frac{1}{\alpha} h(T[u], R)\nabla T[u] \tag{8}$$

where $\nabla T[u]$ is the gradient of the template image in every iteration. The exact form of $h(\cdot,\cdot)$ for three (similarity) measures, i.e. SSD, CR and MI, are given in [9]. The nodal forces computed from (8) are applied to the dynamic model in (7) and the resulting displacement field is used to interpolate a new 3D template image. This new template image is used to calculate the nodal forces in the next iteration. The process continues until the solution converges to an equilibrium.

2.4 2D-3D Deformable Registration

This section presents a method for deformable registration of a 3D preoperative (undeformed) MR image to one or more 2D intraoperative (deformed) MR images. For any 2D slice of the intraoperative image data, the corresponding 2D slice of the preoperative image data is found based on the position and the orientation of the virtual 2D iMRI plane. We employ a method similar to that of Section 2.3 to find the displacement of voxel points to match these two slices. Hence, for every point of the regular grid y, its displacement $dy = -\lambda h(T[u(y)], R)\nabla T[u(y)]$ where $\lambda \in \Re_+$ is a constant gain. It should be noted that dy is a 3×1 vector, i.e. the computed displacement is not restricted to the plane of the 2D slice. The shape functions of the elements relate the displacements of regular grid points of the 2D slice to those of the nodal points of the FE mesh in which these regular points reside. This essentially constitutes the process measurement in our filtering approach to the deformation estimation.

The discretized version of the dynamic deformation model in (7) can be written in the following general state-space form:

$$x(k+1) = Ax(k) + Gf \qquad z(k) = Hx(k) \tag{9}$$

where x is the vector of system states, i.e., nodal displacements and velocities, and $z(k)$ is the vector of measurements, i.e., the computed displacements at the regular 2D grid points. Moreover, A, G, and H are the state transition matrix, the model input matrix and the model output matrix, respectively. Having the model (9) and the measurements, one can iteratively estimate the system states using a Kalman-type filtering process [10]. The deformed 3D intraoperative image can then be obtained by an interpolation of the deformed FE mesh.

3 Experiments and Results

A triple modality biopsy training breast phantom (CIRS model 051) is used for obtaining the experimental data. A 3D volume high-resolution ($512\times512\times136$) set of MR images has been taken from the undeformed phantom using a GE 3T Signa MRI machine. Also, two sets of 3D high-resolution ($512\times512\times136$) and low-resolution ($64\times64\times32$) are taken from the deformed phantom. Fig. 1 shows the apparatus developed for deforming the phantom and two sample images taken in undeformed and deformed states. The device in Fig. 1(a) is made of plexiglass and is MR compatible. Four capsules of vitamin E are attached to the device as landmarks to rigidly register the reference coordinate frames of the deformed and undeformed image data sets. An isotropic linear elastic deformation model with the Young's elasticity modulus $E = 10^5$ and Poisson's ratio $\nu = 0.45$ has been used in the experiments. Since the deformation model merely acts a constraint in the registration process, these parameters could essentially be tuned by the user to achieve a desirable outcome. A cubic mesh of tetrahedral elements encompassing the whole volume of the deformed and undeformed data is created using the COMSOL Software. Fig. 2 depicts the volumetric mesh of

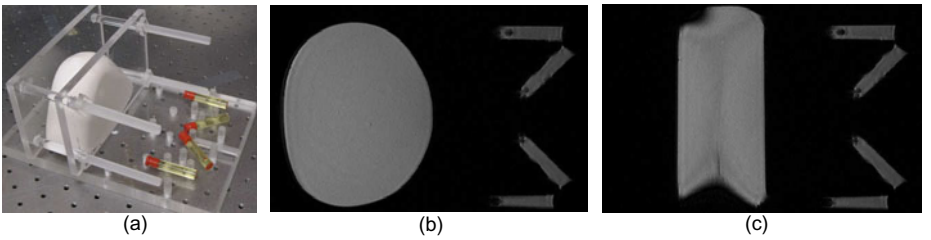


Fig. 1. The experimental setup and sample images: (a) the apparatus used for deforming the breast phantom, (b) image of undeformed phantom, (c) image of deformed phantom

7502 elements and 1601 nodal points before (a) and after (b) deformation. It is clear from this figure that the mesh has to translate and deform in order to cope with the large deformation in the phantom. It should be noted that our approach requires no segmentation of the preoperative image since it does not use a specific geometry for the mesh and applies forces to all nodes.

The evolution of three different distance (similarity) measures in the iterative registration of 3D high-resolution to low-resolution are shown in Fig. 2(c). For this experiment, the registration is done based on the SSD measure, and CR and MI are computed and used for comparison in each iteration. It is evident that the algorithm converges to a solution after about 15 iterations. It is also shown in Fig. 3 that registering the 3D high-resolution preoperative image data (a) to the 3D low-resolution intraoperative image data (b) produces a deformed preoperative image (c) which is very similar to the actual high-resolution intraoperative image (d). A quantitative comparison of the three distance (similarity) measures for three different measures based registration is given in Table 1. For a better comparison, normalized mutual information (NMI) is computed instead of MI because $0 \leq \text{NMI} \leq 2$ for any two sets of image data. The optimal values of the image similarity metrics are zero for SSD, 1 for CR, and 2 for NMI. In Table 1 different distance (similarity) measures for the

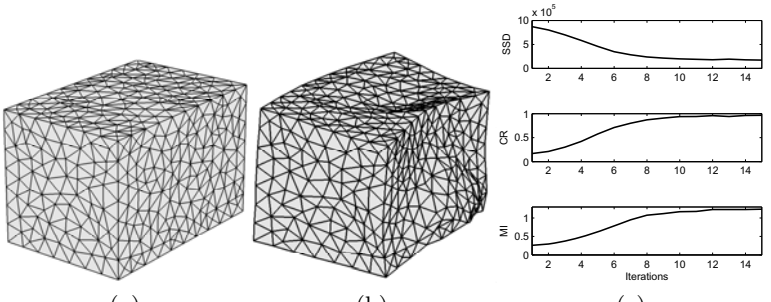


Fig. 2. (a) undeformed mesh, (b) deformed mesh, (c) distance (similarity) measures

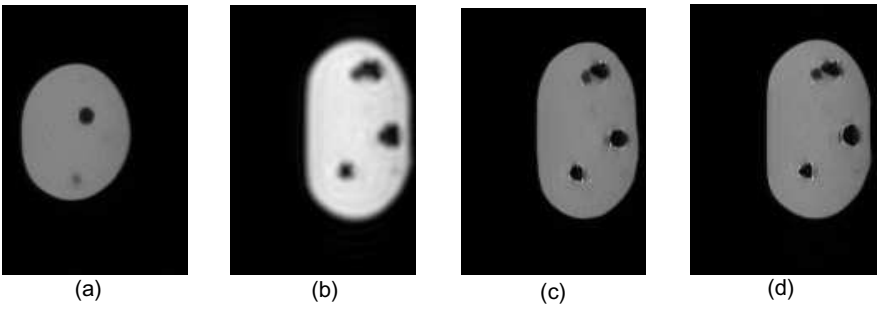


Fig. 3. The x-y view of the (a) preoperative images (template data), (b) low-resolution intraoperative images (reference data), (c) deformed preoperative images after registration, (d) actual high-resolution intraoperative image

Table 1. Distance (similarity) measures for images with (0-255) gray scale values

	3D volumetric registration			2D-3D registration		
Registration	SSD	CR	NMI	SSD	CR	NMI
SSD based	139×10^6	0.941	1.151	167×10^6	0.910	1.061
CR based	141×10^6	0.956	1.159	172×10^6	0.918	1.078
MI based	152×10^6	0.923	1.143	181×10^6	0.882	0.970

case of 2D-3D registration are also given. These measures are computed between the 3D volume of the actual high resolution intraoperative images and the 3D volume of deformed preoperative images obtained through registration. In this case, 15 slices of the 3D volume from various orientations are used and for each slice the iterative Kalman-type filter is executed for 10 time steps.

Table 1 shows that CR-based registration yields the most optimal values of the three metrics, especially in volumetric low-high registration. One reason for the poor performance of MI-based registration may be the rather low resolution images used for registration and the highly nonlinear nature of this similarity measure. Moreover, the results for 3D volumetric registration is better than 2D-3D registration. The 2D-3D method could be improved using better estimation methods and more 2D slices over time.

4 Discussion and Conclusion

A model-based deformable image registration method has been proposed to register high-resolution 3D volume of preoperative MR images to either low-resolution 3D volume or 2D sequences of intraoperative MR images. Within a filtering and estimation framework, the approach employs a continuum mechanics based model of deformation and similarity (distance) measures such as SSD, CR, and MI to non-rigidly register the images. The method converges in few iterations in the presence of large deformations in the experiments with a breast

biopsy phantom. The registration algorithm is based on voxel intensities and requires no feature extraction or image segmentation.

The proposed registration method involves tasks such as three-linear interpolation, solving the equilibrium equations in dynamic FE analysis using the Newmark method [8], finding grid points deformation based on the shape function and matrix computations, which are all computationally expensive. However these tasks are highly amenable to parallel computing using graphics processing units (GPUs). We are currently working on the implementation of the algorithms on a GPU. Our preliminary results for SSD based volumetric registration indicate a factor of 60 speedup over an optimized CPU-based implementation. In future we will extend the algorithm for multi-modal registration. In particular, we will investigate 2D-3D US to MR registration for medical diagnostic as well interventional applications.

Acknowledgments. The authors would like to thank Dr. Michael Noseworthy for providing access to the MRI machine, and Dr. Jan Modersitzki and Dr. Shahram Shirani for their comments and feedback on the work.

References

1. Fahlbusch, R.: Development of Intraoperative MRI: A Personal Journey. Itraoperative Imaging 109, Part 1, 9–16 (2011)
2. Gibson, S.F.F., Mirtich, B.: A Survey of Deformable Modeling in Computer Graphics. TR-97-19 (1997), http://www.ncbi.nlm.nih.gov
3. Ferrant, M., Nabavi, A., Macq, B., Jolesz, F.A., Kikinis, R., Warfield, S.K.: Registration of 3-d intraoperative MR images of the brain using a finite-element biomechanical model. IEEE Trans. on Med. Imaging 20(12), 1384–1397 (2001)
4. Azar, F.S., Metaxas, D.N., Schnall, M.D.: A finite element model of the breast for predicting mechanical deformations during biopsy procedures. In: IEEE Work. on Math. Meth. in Biomed. Ima. Anal., pp. 38–45. Hilton Head Island, SC, USA (2000)
5. Brock, K.K., Sharpe, M.B., Dawson, L.A., Kim, S.M., Jaffray, D.A.: Accuracy of finite element model-based multi-organ deformable image registration. Med. Phys. 32(6), 1647–1659 (2005)
6. Zienkewickz, O.C., Taylor, R.L.: The Finite Element Method. McGraw Hill, New York (1987)
7. Ferrant, M., Warfield, S.K., Guttmann, C.R.G., Mulkern, R.V., Jolesz, F.A., Kikinis, R.: 3D Image Matching Using a Finite Element Based Elastic Deformation Model. In: Taylor, C., Colchester, A. (eds.) MICCAI 1999. LNCS, vol. 1679, pp. 202–209. Springer, Heidelberg (1999)
8. Bathe, K.J.: Finite Element Procedures. Prentice Hall, Englewood Cliffs (1996)
9. Chefd'Hotel, C., Hermosillo, G., Faugeras, O.: A Variational Approach to Multi-Modal Image Matching. In: IEEE Wsh. on Var. Le. Set Meth., Canada, pp. 21–28 (2001)
10. Welch, G., Bishop, G.: An Introduction to the Kalman Filter. Uni. of North Carolina, TR-95-041 (2006)

Using High Resolution Cardiac CT Data to Model and Visualize Patient-Specific Interactions between Trabeculae and Blood Flow

Scott Kulp[1], Mingchen Gao[1], Shaoting Zhang[1],
Zhen Qian[2], Szilard Voros[2], Dimitris Metaxas[1], and Leon Axel[3]

[1] CBIM Center, Rutgers University, Piscataway, NJ, 08550, USA
[2] Piedmont Heart Institute, Atlanta, GA, 30309, USA
[3] New York University, 660 First Avenue, New York, NY, 10016, USA

Abstract. In this paper, we present a method to simulate and visualize blood flow through the human heart, using the reconstructed 4D motion of the endocardial surface of the left ventricle as boundary conditions. The reconstruction captures the motion of the full 3D surfaces of the complex features, such as the papillary muscles and the ventricular trabeculae. We use visualizations of the flow field to view the interactions between the blood and the trabeculae in far more detail than has been achieved previously, which promises to give a better understanding of cardiac flow. Finally, we use our simulation results to compare the blood flow within one healthy heart and two diseased hearts.

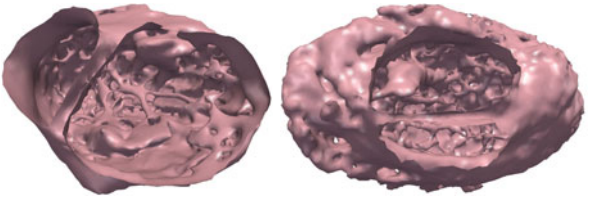
1 Introduction

Following a heart attack or the development of some cardiovascular diseases, the movement of the heart walls during the cardiac cycle may change. This affects the motion of blood through the heart, potentially leading to an increased risk of thrombus. While Doppler ultrasound and MRI can be used to monitor valvular blood flow, the image resolutions are low and they cannot capture the interactions between the highly complex heart wall and the blood flow. For this reason, with the rapid development of high-resolution cardiac CT, patient-specific blood flow simulation can provide a useful tool for the study of cardiac blood flow.

Recently, Mihalef et al. [9] used smoothed 4D CT data to simulate left ventricular blood flow, and compared the flow through the aortic valve in a healthy heart and two diseased hearts. However, the models derived from CT data in [9] were too highly smoothed to capture the local structural details and were not useful for understanding the true interactions between the blood flow and the walls.

Later, in [7], more accurate heart models were achieved by generating a mesh from high-resolution CT data at mid-diastole. Then, motion was transferred to this model from the smooth mesh motion obtained from the same CT data to create the animation. This allowed for more realistic features to be present on the

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 468–475, 2011.



(a) (b)

Fig. 1. Meshes reconstructed from CT data (valves removed). (a) Healthy heart (b) Diseased heart.

heart walls in the simulation, including the papillary muscles and some trabeculae. However, while this approach was an improvement from the smooth-wall assumption, the trabeculae were missing details and did not move accurately.

Earlier work in blood flow simulation used less refined models. For example, [5] was the first to extract boundaries from MRI data to perform patient-specific blood flow simulations. Later, [8] used simple models of the left side of the heart, with smooth ventricular walls, and imposed boundary conditions in the valve regions. In 2010, [6] developed a technique to drive the deformation of a smoothed left ventricle by implicitly coupled fluid motion.

In this paper, we use an improved method of creating the mesh to capture these smaller details and generate a more accurate simulation. To the best of our knowledge, we are able to visualize blood flow in unprecedented detail.

2 Data Acquisition

The CT images were acquired on a 320-MSCT scanner (Toshiba Aquilion ONE, Toshiba Medical Systems Corporation) using contrast agent. This advanced diagnostic imaging system is a dynamic volume CT scanner that captures a whole-heart scan in a single rotation, and achieves an isotropic $0.3mm$ volumetric resolution. A conventional contrast-enhanced CT angiography protocol was adapted to acquire the CT data in this work. After the intravenous injection of contrast agent, the 3D+time CT data were acquired in a single heart beat cycle when the contrast agent was circulated to the left ventricle and aorta. After acquisition, 3D images were reconstructed at 10 time phases in between the R-to-R waves using ECG gating. The acquired isotropic data had an in-plane dimension of 512 by 512 pixels.

The detailed cardiac shape features can be used as the boundary conditions and incorporated in a fluid simulator to derive the hemodynamics throughout the whole heart cycle. Our goal in defining these boundary conditions is to capture the fine detail structures of the myocardium, as well as the one-to-one vertex correspondence between frames, which is required in the fluid simulation. There has been much recent work in cardiac reconstruction, such as [11], who combined high-resolution MRI images with serial histological sectioning data to build histo-anatomically detailed individualized cardiac models to investigate

cardiac function. In this work, we use the techniques described in [3]. Here, snake based semi-automatic segmentation is used to acquire the initial segmentation from high resolution CT data for an initial (3D) frame of data. The initial mesh is generated as an isosurface of the segmentation, which is deformed to match the shape of the heart in each consecutive frame. also during the deformation, we achieve the necessary one-to-one correspondence between frames.

The aortic and mitral valves are thin and move fast, and so the CT data is not currently able to adequately capture these details. We add 3D models of the valves created from ultrasound data to each mesh in the sequence, and open and close the valves at the appropriate time steps.

Reconstruction results for a healthy and a diseased heart can be seen in Figure 1. Note the high level of structural detail at the apex. To the best of our knowledge, this has never been simulated before.

3 Fluid Simulation

The motion of an incompressible fluid is governed by the laws of conservation of momentum and mass, modeled by the Navier-Stokes (NS) equations:

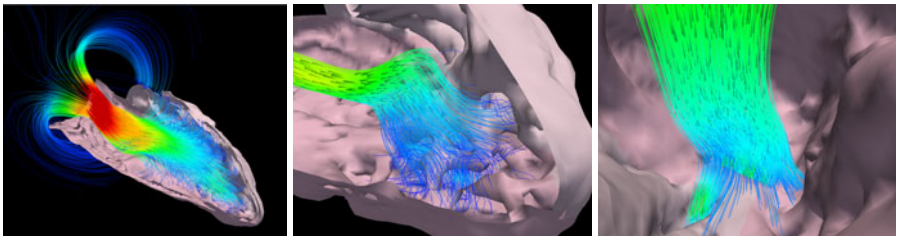
$$\rho(\tfrac{\partial u}{\partial t} + u \cdot \nabla u) = -\nabla P + \mu \nabla^2 u,$$
$$\nabla \cdot u = 0.$$

Here, ρ is the fluid density, u is the 3D velocity vector field, P is the pressure field, and μ is the coefficient of viscosity. We seek to solve these equations for velocity and pressure.

Foster and Metaxas [2] were the first to develop a fast method of solving the NS equations for graphics applications by applying a staggered grid across the domain and explicitly solving for velocity at the cell faces. They then used SOR to solve for pressure and correct the velocities to maintain incompressibility.

Our fluid-solid interaction system uses a "boundary immersed in a Cartesian grid formulation", allowing for an easy treatment of complex moving geometries embedded in the computational domain. Recent work that employs such a formulation is [13], which applies the formulation of [12] to both graphics and medical simulations. Recently, [1] implemented the approach of [4] to obtain a system that can efficiently deal with complex geometric data, such as a system of blood vessels.

The heart models used here are embedded in a computational mesh of 100^3 cells on which the full NS equations are solved using FDM. The blood is modeled as a Newtonian fluid, with viscosity of 4mPa$\cdot$ s and density of 1050kg/m^3, which are physiologically accepted values for normal human blood[9]. The heart model is given to the solver as a set of meshes with point correspondences, which allows for easy interpolation and also obtaining the velocity of the heart mesh at every point in time. Our system represents the 3D meshes as a Marker Level Set(MLS) [10], where markers are placed on the boundary and are used to correct the level set at every time step. Since markers are only placed on the surface, MLS has been proven to be more efficient and more accurate for complex boundaries.



(a) (b) (c)

Fig. 2. Visualization of streamlines within the healthy heart. (a) Streamlines of cardiac blood flow during diastole. (b) Blood flow near apex during diastole. (c) Blood flow during systole at the apex, against the trabeculae.

The MLS and its velocity are rasterized onto the Eulerian grid and are used to impose the appropriate boundary conditions in the fluid solver. A simulation of two complete cardiac cycles takes about four days to complete on a machine with a Core 2 Quad processor and 8GB of RAM.

4 Visualizations

With the fluid velocity fields and level sets generated for each time step, we use Paraview to visualize the simulations. We analyzed a healthy heart and two diseased hearts, and we describe below our visualization methods and our results.

Blood Flow Velocity. We performed a visualization of the velocity field within the heart, as seen in Figure 3. The velocity of the blood at a given point is represented by a cone pointed in the direction of the flow. The size of cone increases linearly as the velocity increases. We also adjust the color of a cone by first setting its hue to 160 (blue), and then linearly lowering this value to a minimum of 0 (red) as velocity increases. The magnitude of fluid velocity ranges from 0-.9 m/s.

Streamline visualizations are shown in Figure 2. The color at a point within a streamline is chosen in the same way as the cones described above. In order to disambiguate direction, we add cones that point in the direction of flow

4.1 Blood Residence Time

In addition to the blood flow velocities, we wish to visualize the residence time of blood within the heart. By doing so, we can quantitatively determine regions of the heart that are at greater risk of thrombus, as slower flows are known to be a significant factor predisposing to thrombus formation.

At the initial time step, ten thousand particles are generated randomly within the heart. At the beginning of each time step, new particles are generated at the valves, allowing fresh blood particles to enter the heart during diastole. Each new particle has an initial age of zero, and this age is incremented at every time step.

At each consecutive time step, we determine a particle's velocity by interpolation, given the fluid velocities at the center of each cell. Each particle's new position is calculated using Euler time integration. Then, any particle in a cell exterior to the heart is removed from the system, and the average particle residence time within each cell can then be easily determined. We run this for four cardiac cycles and create volumetric visualizations, as seen in Figure 4. Here, blue represent regions in which average residency is less than 1 cardiac cycle, green-yellow represents 1-3 cardiac cycles, and red represents 3-4 cycles.

We can also take advantage of these particles in validation of our simulation, by computing an estimated ejection fraction. During systole, we know exactly how many particles there originally existed in the system, and how many are being expelled at each time step. To estimate the ejection fraction, we simply divide the total number of deleted particles by the original number of particles.

5 Discussion

The streamline visualizations provide detailed information on the trabeculae-blood interaction. Figure 2(b), taken during diastole, demonstrates how the complex surface causes the flow to move through and around the empty spaces between the trabeculae. Then, in Figure 2(c), during systole, we see another example of how the blood is forcefully expelled out of the spaces between the trabeculae, rather than simply flowing directly towards the aortic valve as older methods with simpler meshes have suggested.

The simulation and visualization methods are performed described above on three different hearts. The first is a healthy heart with no visible medical problems with an ejection fraction of about 50%. The second is a heart that has simulated hypokinesis, where the motion of the heart walls is decreased at the apex by a maximum of 50%. The third comes from a patient who has post tetralogy of Fallot repair. This heart is known to suffer from right ventricle hypertrophy, significant dyssynchrony in the basal-midseptum of the left ventricle, and a decreased left ventricle ejection fraction of about 30%.

The streamline visualizations provide detailed information on the trabeculae-blood interaction. Figure 2(b), taken during diastole, demonstrates how the complex surface causes the flow to fill the empty spaces between the trabeculae. Then, in Figure 2(c), during systole, we see another example of how the blood is expelled out of the spaces between the trabeculae, rather than simply flowing directly towards the aortic valve as older methods with simpler meshes have suggested.

Validation is a difficult task, as current imaging techniques, such as PC-MRI, are not able to capture flow information at the required level of detail for useful comparison. We performed a partial validation by comparing the estimated ejection fraction to the true ejection fraction. The computed ejection fraction is approximately 45% for the healthy heart, 40% for the hypokinesis heart, and 30% for the dyssynchronous heart. These values for the healthy and dyssynchronous heart are in agreement with the true values, so we have confidence in

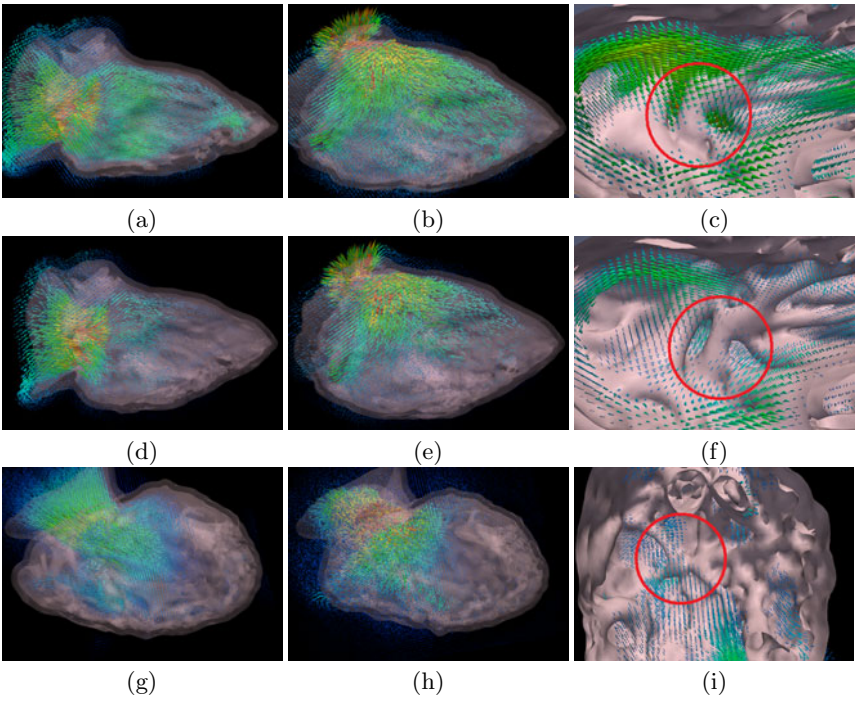
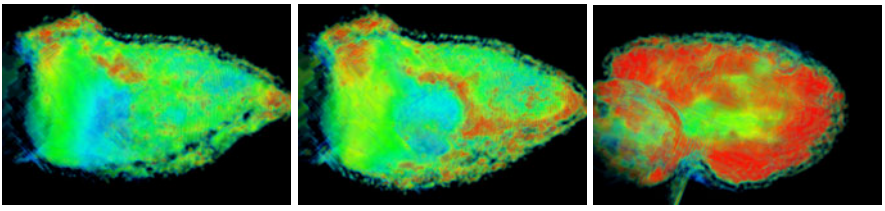


Fig. 3. Velocity fields at various time steps for three different hearts. Top row: Healthy Heart, Middle row: Hypokinetic heart, Bottom row: Dyssynchronous heart. Left column: Diastole, Middle column: Systole, Right column: Velocity field at trabeculae during systole.

the rest of our results. Performing similar validation techniques to a smoothed healthy heart model, we computed an ejection fraction of about 40%, slightly lower than that of our complex model. However, it may not be especially useful to compare the accuracy of different modeling methods using this approach, as the ejection fraction does not give information about the flow local to the apex, the region of primary interest.

Velocity field visualizations are illustrated in Figure 3. We can see that in the healthy heart, the inflow during diastole is significant and fairly uniformly distributed, circulating blood throughout the heart. During systole, the velocity field throughout the heart remains high, and fluid in the apex moves toward the valves. In Figure 3(c), we see more detail of the interactions between blood flow and the trabeculae, as the blood is visibly expelled from these regions. However, in the heart suffering from hypokinesis, we find that the velocity field is much weaker toward the apex during both diastole and systole. In Figure 3(f), we also see that the trabeculae are no longer adequately expelling blood as they do in the healthy heart case. We also see in Figure 3(g)-(i) that the flow patterns in the heart with dyssynchronous heart wall movement appears non-normal, with overall lower velocities and even less fluid being pushed out from the trabeculae.



(a) (b) (c)

Fig. 4. Visualization of average particle residence time. Colors closer to red represent longer average residence time. (a) Healthy Heart (b) Heart with Hypokinesis (c) Heart with dyssynchronous wall movement.

We then compare the visualizations of the average particle residence times for each of the three simulations, as seen in Figure 4. Each of these images were made at the same time step, at the start of systole, after four cardiac cycles. We find that in Figure 4(a), in the healthy heart, nearly the entire domain contains blood with average residence time of less than three cycles, suggesting that the blood is not remaining stagnant, and turning over well between cardiac cycles. In contrast, Figure 4(b) shows that in the heart suffering from hypokinesis, the average residence time is significantly higher near the walls, particularly near the hypokinetic apex. Finally, in Figure 4(c), we find that a very significant region of the blood has a long residence time, suggesting that due to the low ejection fraction and relatively low fluid velocities, blood is not being adequately circulated and thus is remaining stagnant near the walls, again, particularly toward the apex of the heart.

6 Conclusions

In this paper, we have described our new framework to generate detailed mesh sequences from CT data, and used them to run patient-specific blood flow simulations. We then created several visualizations to reveal the interactions between the complex trabeculae of the heart wall and the blood, which has never been possible before, and used them to compare the flow fields between a healthy heart and two diseased hearts, which would potentially be extremely useful to doctors to help in diagnosis and treatment plans. This is the first time that intracardiac blood flow fields and their interaction with the heart wall have been investigated at this level of resolution.

Acknowledgments. This material is based upon work supported by the U.S. Department of Homeland Security under Grant Award Number 2007-ST-104-000006. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Department of Homeland Security.

References

1. de Zelicourt, D., Ge, L., Wang, C., Sotiropoulos, F., Gilmanov, A., Yoganathan, A.: Flow simulations in arbitrarily complex cardiovascular anatomies - an unstructured cartesian grid approach. Computers & Fluids 38(9), 1749–1762 (2009)
2. Foster, N., Metaxas, D.: Realistic animation of liquids. Graph. Models Image Process. 58, 471–483 (1996)
3. Gao, M., Huang, J., Zhang, S., Qian, Z., Voros, S., Metaxas, D., Axel, L.: 4D cardiac reconstruction using high resolution ct images. In: Metaxas, D., Axel, L. (eds.) FIMH (2011)
4. Gilmanov, A., Sotiropoulos, F.: A hybrid cartesian/immersed boundary method for simulating flows with 3D, geometrically complex, moving bodies. J. Comput. Phys. 207(2), 457–492 (2005)
5. Jones, T.N., Metaxas, D.: Patient-specific analysis of left ventricular blood flow. In: Wells, W.M., Colchester, A.C.F., Delp, S.L. (eds.) MICCAI 1998. LNCS, vol. 1496, pp. 156–166. Springer, Heidelberg (1998)
6. Krittian, S., Janoske, U., Oertel, H., Bhlke, T.: Partitioned fluid-solid coupling for cardiovascular blood flow. Annals of Biomedical Engineering 38, 1426–1441 (2010)
7. Kulp, S., Metaxas, D., Qian, Z., Voros, S., Axel, L., Mihalef, V.: Patient-specific modeling and visualization of blood flow through the heart. In: Wright, S., Pan, X., Liebling, M. (eds.) ISBI (2011)
8. Long, Q., Merrifield, R., Yang, G.Z., Xu, X.Y., Kilner, P.J., Firmin, D.N.: The influence of inflow boundary conditions on intra left ventricle flow predictions. Journal of Biomechanical Engineering 125(6), 922–927 (2003)
9. Mihalef, V., Ionasec, R., Wang, Y., Zheng, Y., Georgescu, B., Comaniciu, D.: Patient-specific modeling of left heart anatomy, dynamics and hemodynamics from high resolution 4d CT. In: Niessen, W., Meijering, E. (eds.) ISBI, pp. 504–507 (2010)
10. Mihalef, V., Metaxas, D., Sussman, M.: Textured liquids based on the marker level set. Comput. Graph. Forum 26(3), 457–466 (2007)
11. Plank, G., Burton, R., Hales, P., Bishop, M., Mansoori, T., Burton, M., Garny, A., Prassl, A., Bollensdorff, C., Mason, F., Mahmood, F., Rodriguez, B., Grau, V., Schneider, J., Gavaghan, D., Kohl, P.: Generation of histo-anatomically representative models of the individual heart: tools and application. In: Phil. Trans. Roy. Soc. (2009)
12. Sussman, M.: A parallelized, adaptive algorithm for multiphase flows in general geometries. Comput. Struct. 83(6-7), 435–444 (2005)
13. Yokoi, K., Xiao, F., Liu, H., Fukasaku, K.: Three-dimensional numerical simulation of flows with complex geometries in a regular cartesian grid and its application to blood flow in cerebral artery with multiple aneurysms. Journal of Computational Physics 202(1), 1–19 (2005)

Motion Correction and Parameter Estimation in dceMRI Sequences: Application to Colorectal Cancer

Manav Bhushan[1,2,⋆], Julia A. Schnabel[1], Laurent Risser[1], Mattias P. Heinrich[1,2], J. Michael Brady[3], and Mark Jenkinson[2]

[1] Institute of Biomedical Engineering, Oxford University, UK
[2] Oxford University Centre for Functional MRI of the Brain, UK
[3] Department of Radiation Oncology and Biology, Oxford University, UK

Abstract. We present a novel Bayesian framework for non-rigid motion correction and pharmacokinetic parameter estimation in dceMRI sequences which incorporates a physiological image formation model into the similarity measure used for motion correction. The similarity measure is based on the maximization of the joint posterior probability of the transformations which need to be applied to each image in the dataset to bring all images into alignment, and the physiological parameters which best explain the data. The deformation framework used to deform each image is based on the diffeomorphic logDemons algorithm. We then use this method to co-register images from simulated and real dceMRI datasets and show that the method leads to an improvement in the estimation of physiological parameters as well as improved alignment of the images.

1 Introduction

Colorectal cancer is the third most common cancer and the fourth largest cause of cancer deaths worldwide, with an estimated 1.2 million cases and 609,000 deaths in 2008 alone [5]. The standard protocol for treating people with colorectal cancer is to give the patient 6-12 weekly cycles of chemoradiotherapy and then to surgically remove the tumour. Roughly 30% patients do not respond to chemoradiotherapy, and hence should have gone straight to surgery (non-responders), whereas 15-20% respond completely and still undergo unnecessary surgery (responders). Although there is currently a lack of any objective mechanism of discriminating between these two groups at an early stage in the treatment, dynamic contrast-enhanced MRI (dceMRI) has recently emerged as a promising non-invasive imaging technique that could help in making these clinical decisions [11].

Anatomical MRI is used widely in oncology since it shows good contrast between soft tissues, and because it does not involve any harmful radiation.

⋆ We would like to thank EPSRC and Cancer Research UK for funding this work within the Oxford Cancer Imaging Centre. JAS, LR also acknowledge funding from EPSRC EP/H050892/1 and MJ acknowledges funding from BBSRC.

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 476–483, 2011.

In dceMRI, the patient is injected with a paramagnetic contrast agent (CA) and then imaged every 10-12 seconds for 5-6 minutes after injection. The data obtained can be used to calculate physiologically relevant parameter maps by fitting an appropriate pharmacokinetic (PK) model to the signal-intensity curve obtained at each voxel, and these maps can then be used to assess tissue perfusion and heterogeneity. However, the biggest challenge in trying to calculate these PK parameter values for each voxel is posed by patient motion during the scan, which causes a voxel in the image to correspond to different anatomical locations at different time points. This problem is more pronounced in the case of colorectal images, since movement in this area cannot be constrained.

The problem of PK-parameter estimation for dceMRI images has been addressed previously in [4], [7], [3] and [12]. In [7] the authors focussed on estimation of PK-parameters for each voxel using a Bayesian model, whereas in [3] and [12] motion correction and parameter estimation were tackled in a sequential manner. In [3] the authors incorporate a PK model into the similarity measure used for registration of images, but this model only allows for simple translations of the image and is based on curve-fitting at each iteration, which is computationally very expensive. In [12], the images are registered using mutual information as a similarity measure and B-Splines as a deformation framework. The PK-parameters are then estimated by fitting curves to the data for each voxel. In this paper we have addressed both problems in a unified probabilistic framework. In Section 2 we describe our algorithm in detail. In Section 3 we show results on simulated and real data, and finally conclude in Section 4.

2 Methods

2.1 Similarity Measure

The main contribution in this paper is to derive a Bayesian similarity measure for simultaneous motion correction and PK-parameter estimation in dceMRI scans. It has been shown in [6] that most of the conventional similarity measures used in image registration can be derived in a Bayesian manner if we assume certain image formation models. We have followed the same generalized approach to derive a similarity measure for motion-correction in dceMRI sequences by assuming an image formation model based on a PK model. The key assumption that we make in order to define our Bayesian similarity measure is that the 'true' uncorrupted image X_i (for each time-point $i = 1, ..., t$) is a function of some physiological characteristics of the tissue, i.e. $X_i = f_i(\boldsymbol{\theta})$. The similarity measure used to co-register all the images obtained during the scan is based on maximizing the joint posterior probability of the set of transformations $\boldsymbol{T} = \{T_i\}_{i=1,..,t}$ that need to be applied to the observed images in order to recover the 'ground-truth' images and the set of PK-parameters $\boldsymbol{\theta} = \{\theta_j\}_{j=1,..,N}$, which provide the best-fit curve to the intensities observed at each voxel. In this paper, we have assumed that the concentration of CA at each voxel is governed by

the Tofts pharmacokinetic model [8] assuming the arterial input function proposed by Weinmann et al. [10], and is given by:

$$C(t) = DK_{ep}V_e \left\{ 3.99 \left(\frac{e^{-K_{ep}t} - e^{-0.144t}}{0.144 - K_{ep}} \right) + 4.78 \left(\frac{e^{-K_{ep}t} - e^{-0.0111t}}{0.0111 - K_{ep}} \right) \right\} \quad (1)$$

Here $C(t)$ is the concentration of CA at time t, K_{ep} and V_e are the two parameters of the model (per voxel), and D is the dosage of the CA. We have chosen this because it is a simple yet widely accepted PK model that has physiologically relevant model parameters. However, it is important to remember that the PK model is simply one component of our algorithm, and could be replaced with any other PK model if desired.

The MR intensity observed at time t is a function of $C(t)$, the MR intensity before injection of CA (i.e. the pre-contrast image X_0), the T1-relaxation time of the tissue, and the MRI acquisition parameters. Previously, the concentration of CA and the signal intensity were thought to be linearly related, but recent studies have shown that this relationship is quite complex and non-linear [1]. In our implementation, we have incorporated this non-linear relationship between the CA concentration and the MR intensity for greater accuracy.

In our model, we assume that the set of images actually observed, $\boldsymbol{Y} = [Y_i]_{\,i=1,..,t}$ is obtained when each original image X_i is deformed (due to patient-motion) and then subjected to some noise process, i.e.

$$Y_i = T_i\left(X_i\right) + \epsilon_i \quad (2)$$

Here, ϵ_i is a Gaussian noise-process with 0 mean and standard deviation σ_i. In this paper, we assume that the standard deviation for the noise process is constant for all time-points, i.e. $\forall i,\ \epsilon_i \sim N\left(0, \sigma\right)$. The physiological parameters of interest at each voxel are given by $\theta = [K_{ep}, V_e]$ and the intensity of voxel j at time-point i can be written as $f_i(\theta_j)$. It is worthwhile noting that the intensity in a voxel at time t also depends on the intensity at that voxel in the pre-contrast image X_0, but this dependence has been left implicit in our notation. The entire ground-truth image X_i can be expressed as $f_i(\boldsymbol{\theta})$, where $\boldsymbol{\theta} = \{\theta_j\}_{j=1,..,N}$ and we can re-write our image formation model as $\boldsymbol{Y} = \boldsymbol{T}\left(\boldsymbol{f}(\boldsymbol{\theta})\right) + \boldsymbol{\epsilon}$, where each term is actually a t-dimensional vector, and each element of this vector is an image with N voxels. In this scenario, the likelihood term becomes:

$$P\left(\boldsymbol{Y} \mid X_0, \boldsymbol{T}, \boldsymbol{\theta}, \sigma\right) = P\left(\epsilon = \left\|\boldsymbol{f}(\boldsymbol{\theta}) - \boldsymbol{T}^{-1}\left(\boldsymbol{Y}\right)\right\|\right) \quad (3)$$

$$= \frac{1}{\left(2\pi\sigma^2\right)^{N/2}} \exp\left(\frac{-\left\|\boldsymbol{f}(\boldsymbol{\theta}) - \boldsymbol{T}^{-1}\left(\boldsymbol{Y}\right)\right\|^2}{2\sigma^2} \right) \quad (4)$$

Here X_0 is the true pre-contrast image (which is known to us) and σ is the unknown noise parameter. We are interested in estimating the set of transformations $\hat{\boldsymbol{T}}$, and PK parameters $\hat{\boldsymbol{\theta}}$ which maximize the joint posterior probability:

$$P(\boldsymbol{\theta}, \boldsymbol{T} \mid X_0, \boldsymbol{Y}, \sigma) \propto P\left(\boldsymbol{Y} \mid X_0, \boldsymbol{T}, \boldsymbol{\theta}, \sigma\right) P(\boldsymbol{\theta}) P(\boldsymbol{T}) \quad (5)$$

In this paper, we assume that the pre-contrast image X_0, the set of transformations $\boldsymbol{T}$, the physiological parameter $\boldsymbol{\theta}$ and the noise parameter σ are all independent of each other and that the prior distributions on these terms are uninformative. However, our framework allows us to incorporate informative priors on any of these terms if required. Thus, the basic steps of our algorithm are:

- Step 1: Choose initial values for $\boldsymbol{\theta}$ and $\boldsymbol{T}$ and calculate $P(\boldsymbol{\theta}, \boldsymbol{T} \mid X_0, \boldsymbol{Y}, \sigma)$.
- Step 2: Calculate $\frac{\partial}{\partial \boldsymbol{T}} \left(P(\boldsymbol{\theta}, \boldsymbol{T} \mid X_0, \boldsymbol{Y}, \sigma)\right)$, $\frac{\partial}{\partial \boldsymbol{\theta}} \left(P(\boldsymbol{\theta}, \boldsymbol{T} \mid X_0, \boldsymbol{Y}, \sigma)\right)$ using the current estimates of $\boldsymbol{\theta}$ and $\boldsymbol{T}$.
- Step 3: Use the gradients calculated in Step 2 to update the values of $\boldsymbol{\theta}$ and $\boldsymbol{T}$ (simultaneously) so that $P(\boldsymbol{\theta}, \boldsymbol{T} \mid X_0, \boldsymbol{Y}, \sigma)$ increases.
- Step 4: Repeat Steps 2 and 3 using the new estimates of $\boldsymbol{\theta}$, $\boldsymbol{T}$ and $P(\boldsymbol{\theta}, \boldsymbol{T} \mid X_0, \boldsymbol{Y}, \sigma)$ until convergence.

In Step 1, the initial value of $\boldsymbol{\theta}$ is chosen by fitting a curve through the initial data for each voxel, and $\boldsymbol{T}$ is initialized as the identity transformation. Since the choice of σ only scales the gradients, we set it to 1. The main step in this procedure is to calculate the analytic derivative of the joint posterior probability with respect to the transformation parameters and the PK-parameters. We then use these derivatives in a Gauss-Newton optimization framework to find:

$$[\hat{\theta}, \hat{T}] = \arg\max_{\boldsymbol{\theta}, \boldsymbol{T}} P(\boldsymbol{\theta}, \boldsymbol{T} \mid X_0, \boldsymbol{Y}, \sigma) \tag{6}$$

2.2 Diffeomorphic Registration Framework

Since our objective is to correct for patient motion during a single scan, it would be natural to assume that there should be no change in topology when we apply the deformations. In other words, we would expect all deformations to be invertible and smooth. For these reasons, we have chosen to use the diffeomorphic deformation framework based on the LogDemons approach followed in [9], where the transformation that needs to be applied to each image is parametrized by a stationary velocity field instead of a displacement field. In other words, there is a 3D velocity vector associated to each voxel, and each transformation T_i is parametrized by the matrices $[U_i, V_i, W_i]$, where U, V and W are the components of the velocity field along the X, Y and Z axis respectively. In our implementation, the update to the velocity field at a particular time point i is taken to be $dU_i = \frac{\partial}{\partial U_i} \left(P(\boldsymbol{\theta}, \boldsymbol{T_i} \mid X_0, Y, \sigma)\right)$, and similarly for V and W. We then find the resulting deformation by dividing the velocity field into small time-steps and integrating over it for one unit of time, as explained in [9].

3 Results

3.1 Tests on Simulated Data

Our main contribution in the motion-correction and parameter estimation algorithm described above is the Bayesian similarity measure that has been used to

co-register all the images. Thus, we were interested in comparing our method with motion correction algorithms using the same deformation framework, but different similarity measures. For this purpose, we compared our algorithm with the same framework using sum-of-squared differences (SSD), mutual information (MI) and block-wise normalized cross-correlation (NCC) as a similarity measure, as described in [2]. In order to use these other similarity measures within our framework, we replaced the gradient with respect to the transformation T calculated in Step 2 of our algorithm with the gradient of the respective similarity functions, and ignored the gradient with respect to the PK-parameters.

We generated 5 simulated datasets by choosing a small ROI in a pre-contrast image volume from a real dceMRI scan and simulated contrast uptake using the Tofts model and 5 different artificial parameter maps for θ (starting with only one spherical region of high θ values at the centre, and then choosing a larger number of high-θ regions for each new dataset). The pre-contrast ROI volume was of dimensions $100 \times 100 \times 52$, and we used 10 different time-points for each dataset. To simulate motion, we deformed each image independently by taking a grid of dimensions $5 \times 5 \times 4$ and applied a random displacement to each grid point and then upsampled the grid to the image volume. We applied Gaussian smoothing to the final deformation field before applying it to the image with simulated contrast. This is equivalent to using a radial basis function for each image and convolving it with a Gaussian kernel. The parameters for the artificial deformations were chosen to be such that the maximum deformation in the final deformation fields were of the order ± 3.5 voxels (± 5 mm) along each direction, and there were no negative Jacobians.

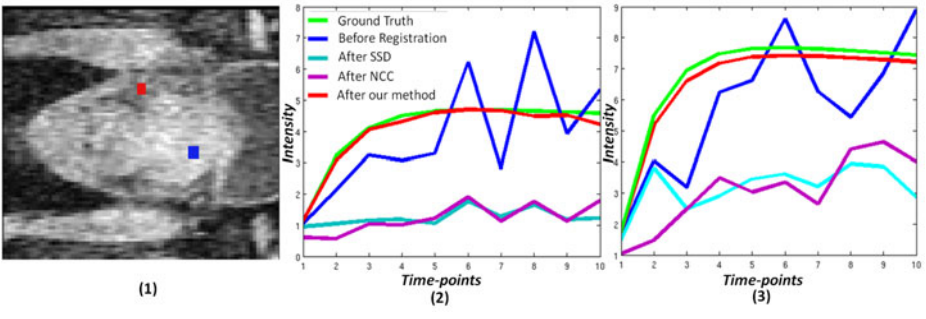


Fig. 1. An axial slice of the 1st time-point image (1) and voxel plots for the locations shown in red (2) and blue (3). The signal-intensity curve in the ground-truth image is shown in green, the one observed before registration is shown in dark blue, the one obtained after SSD registration is shown in light blue, after NCC registration in purple, and after applying our method in red.

Table (1) shows the average results obtained after testing each method on the 5 datasets. In order to test SSD, MI and NCC, we registered all the images in each dataset to the first time-point image (Method A), the 5th time-point

Table 1. Average results on simulated data. BR stand for Before Registration, MSE for Mean Squared Error in intensity values (between the registered images and the original simulated images), and TRE for Target Registration Error (i.e. the average difference between the ground truth deformation field, and the deformation field estimated by the algorithm).

		Method A			Method B			Method C			
Measure	BR	MI	SSD	NCC	MI	SSD	NCC	MI	SSD	NCC	Our Method
MSE	0.159	0.101	0.116	0.062	0.096	0.139	0.061	0.095	0.081	0.062	0.057
Error in K_{ep}	0.462	0.168	0.095	0.046	0.155	0.103	0.043	0.089	0.075	0.039	0.034
Error in V_e	0.116	0.054	0.059	0.021	0.043	0.072	0.021	0.037	0.023	0.019	0.016
TRE	1.933	1.161	1.087	0.765	1.178	1.109	0.729	1.152	0.876	0.743	0.678

image (Method B), and each image to the previous time-point image (Method C). After applying each method, we estimated the PK-parameters for each voxel by fitting curves (using least-squares fitting) in the registered dataset. Thus, the errors in K_{ep} and V_e denote the average absolute difference between parameter values used to generate the dataset, and those estimated from the registered dataset. Our method took 12 mins and NCC 4 mins for the co-registration of all images within each dataset. The standard deviation of the errors in estimation of PK-parameters was about 0.009 (our method) and 0.011 (NCC using Method C) for K_{ep} and 0.0027 (our method) and 0.0043 (NCC using Method C) for V_e. It is worthwhile noting that in our method the PK-parameter values are updated in each iteration and the final values are obtained as outputs at the completion of the registration algorithm. Thus our method does not require curve-fitting (which is computationally expensive) at the end of the registration process.

3.2 Tests on Real Data

After comparing our method with other similarity measures on simulated data, we applied our method to datasets for 6 patients with histologically proven rectal adenocarcinoma (2 responders and 4 non-responders). High spatial resolution 3D T2-weighted images, as well as 3D dceMRI scans were acquired for each of the patients before, and after 5 weeks of chemoradiotherapy. After therapy, tumours were surgically removed and assessed histopathologically to determine the grade. The MRI sequence used to acquire each image during the dceMRI scan was a T1-weighted, gradient-echo, fat-suppressed sequence (LAVA) with TR=4.5ms, TE=2.2ms and flip angle $\alpha = 12$ degrees. The contrast agent (gadobenate dimeglumine, also known as MultiHance) was injected via a peripheral vein after 4 initial acquisitions and MRI volumes (of dimensions $512 \times 512 \times 52$) were acquired every 12-15 seconds for the next 5 minutes for each patient.

We applied our algorithm to the tumour ROI (shown in the inset in Fig. 2) which had dimensions $100 \times 100 \times 52$ from all the images acquired immediately after injection of the CA, and before the second pass of the CA (to ensure that the Tofts model is applicable to the data). We also applied the motion-correction algorithm using NCC (Method C) as a similarity measure (since this

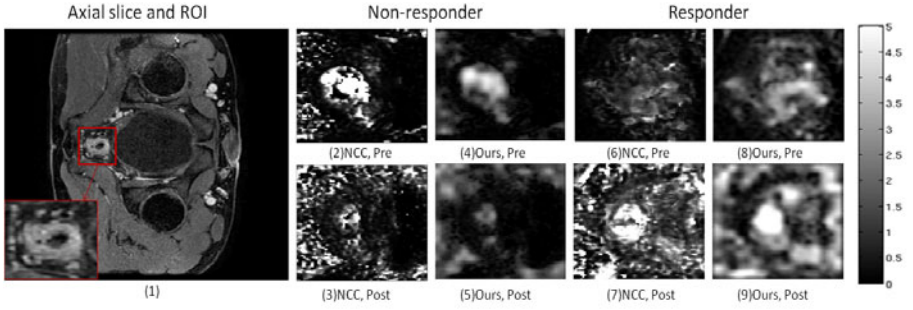


Fig. 2. An axial slice of a post-contrast image volume and the tumour ROI in inset (1). Axial slices of the pre-therapy (2,4) and post-therapy (3,5) K_{ep} maps for a *non-responder* obtained after motion correction using NCC (2,3) and after using our method (4,5). Axial slices for pre-therapy (6,8) and post-therapy (7,9) K_{ep} maps for a *responder* after motion correction using NCC (6,7) and using our method (8,9).

performed best after our method on the simulated data) and then compared the results that we obtained for the pre-therapy and post-therapy scans for each patient. Although the results were qualitative, we found that the parameter maps obtained after motion correction using NCC were generally more noisy, and generally higher than the values estimated using our method (Fig 2). Fig 3 demonstrates that the algorithm using our similarity measure has the potential to distinguish responders from non-responders more clearly than NCC, since we would expect the distribution of K_{ep} values for a non-responder to stay the same, and for a responder to change after treatment.

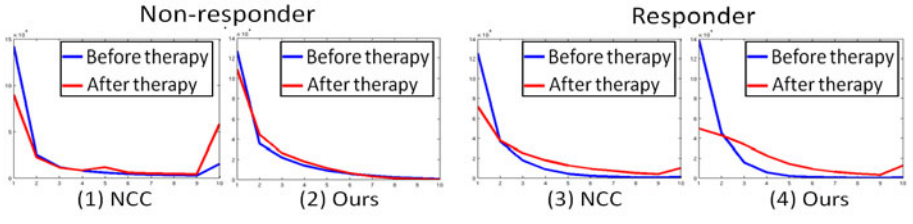


Fig. 3. The histograms (using 10 equally spaced bins) for the K_{ep} values estimated for a non-responder using NCC (1) and our method (2), and for a responder using NCC (3) and our method (4). Blue and red show the pre-therapy and post-therapy distributions respectively.

4 Conclusion

We have developed a Bayesian framework for simultaneous motion-correction and parameter estimation in dceMRI sequences. The derived similarity measure has been compared with other conventional similarity measures used in image

registration and shown to outperform them under four different error metrics. The method has been tested on real data for 6 patients and shown to have the potential for improving the distinction between responders and non-responders. This generic Bayesian framework will also allow us to incorporate other PK-models and obtain a probabilistic quantification of the PK-parameter maps and their uncertainty.

References

1. Armitage, P., Behrenbruch, C., Brady, M., Moore, N.: Extracting and visualizing physiological parameters using dynamic contrast-enhanced magnetic resonance imaging of the breast. Medical Image Analysis 9(4), 315–329 (2005)
2. 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(1), 26 (2008)
3. Buonaccorsi, G.A., Roberts, C., Cheung, S., Watson, Y., Davies, K., Jackson, A., Jayson, G.C., Parker, G.J.M.: Tracer kinetic model-driven registration for dynamic contrast enhanced MRI time series. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3749, pp. 91–98. Springer, Heidelberg (2005)
4. Hayton, P., Brady, M., Tarassenko, L., Moore, N.: Analysis of dynamic MR breast images using a model of contrast enhancement. Medical Image Analysis 1(3), 207–224 (1997)
5. Karsa, L., Lignini, T.A., Patnick, J., Lambert, R., Sauvaget, C.: The dimensions of the CRC problem. Best Pract. Res. Cl. Ga. 24(4), 381–396 (2010)
6. Roche, A., Malandain, G., Ayache, N.: Unifying maximum likelihood approaches in medical image registration. Int. J. Imag. Syst. Tech. 11(1), 71–80 (2000)
7. Schmid, V.J., Whitcher, B., Padhani, A.R., Taylor, N.J., Yang, G.Z.: A Bayesian hierarchical model for the analysis of a longitudinal dynamic contrast-enhanced MRI oncology study. Magn. Reson. Med. 61(1), 163–174 (2009)
8. Tofts, P.S., Kermode, A.G.: Measurement of the blood-brain barrier permeability and leakage space using dynamic MR imaging 1. Fundamental Concepts. MR in Medicine 17(2), 357–367 (1991)
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. Weinmann, H.J., Brasch, R.C., Press, W.R., Wesbey, G.E.: Characteristics of gadolinium-DTPA complex: a potential NMR contrast agent. AJR 142(3), 619 (1984)
11. Zahra, M.A., Hollingsworth, K.G., Sala, E., Lomas, D.J., Tan, L.T.: Dynamic contrast-enhanced MRI as a predictor of tumour response to radiotherapy. The Lancet Oncology 8(1), 63–74 (2007)
12. Zöllner, F.G., Sance, R., Rogelj, P., Ledesma-Carbayo, M.J., Rørvik, J., Santos, A., Lundervold, A.: Assessment of 3D dceMRI of the kidneys using non-rigid image registration and segmentation of voxel time courses. Comput. Med. Imaging Graph 33(3), 171–181 (2009)

Sparse Dose Painting Based on a Dual-Pass Kinetic-Oxygen Mapping of Dynamic PET Images

Kuangyu Shi[1], Sabrina T. Astner[1], Liang Sun[2], Nassir Navab[3], Fridtjof Nüsslin[1], Peter Vaupel[1], and Jan J. Wilkens[1]

[1] Radiation Oncology, Technische Universität München, Munich, Germany
[2] Computer Science and Engineering, Arizona State University, Tempe, USA
[3] Computer Aided Medical Procedures, Technische Universität München, Munich, Germany
shi@lrz.tum.de

Abstract. Development of molecular imaging such as positron-emission tomography (PET) offers an opportunity to optimize radiotherapy treatment planning by conforming the dose distribution to physiological details within tumors, so called dose painting. Quantification of the acquired images and an efficient and practical dose prescription remain two key questions in this field. This paper proposes a novel framework to optimize the dose prescription based on dual-pass modeling of dynamic [^{18}F]FMISO PET images. An optimization algorithm for sparse dose painting (SDP) is developed by minimizing a linear combination of two terms corresponding to the efficiency and total variation of the dose distribution with the constraint of a constant mean dose. Dose efficiency is defined using the linear-quadratic model. The radiosensitivity given by the oxygen tension is estimated using a dual-pass kinetic-oxygen mapping strategy. This is achieved by integrating a realistic [^{18}F]FMISO PET imaging simulation model, which can simulate the distribution of oxygen and tracer under the same tumor microenvironment setting. The algorithm was compared with a typical dose painting by number (DPBN) method in one data set of a patient with head and neck cancer.

Keywords: Dose Painting, Sparse Optimization, Radiation Therapy, Computational Simulation, Positron Emission Tomography.

1 Introduction

Medical imaging plays an important role in radiotherapy including treatment planning [1,2], delivery [3] and monitoring. Recent developments have focused on the incorporation of image derived physiological information into treatment planning [1,2]. Molecular imaging modalities, such as positron-emission tomography (PET), can be applied to improve the therapy outcome by conforming the dose distribution to the heterogeneous tumor microenvironment, which is called dose painting [4]. Inverse planning and intensity-modulated radiotherapy (IMRT) with dynamic multileaf collimators are applied to shape the delivery to match the dose prescription [5]. To prescribe a proper and efficient dose distribution based on the acquired molecular images is one key factor for successful dose painting.

In practice, two strategies have been applied to paint dose values within a tumor: (1) boost escalation, which outlines the resistant regions and escalates uniform doses

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 484–491, 2011.

correspondingly; (2) dose painting by numbers (DPBN), which prescribes the dose according to the radiosensitivity difference of each image voxel [6,7]. Concerning the highly heterogeneous tumor microenvironment [8], DPBN has advantages in conformance to biological details. However, it is often hardly feasible in practice due to the exploded complexities in planning and delivery procedure, which not only increases the burden to patients and machines, but also reduces the freedom of normal tissue sparing. To prescribe a practical dose distribution while keeping the efficiency remains challenging in dose painting. Sparsity regularization is one booming technology to explore the information usage while reducing the data complexity, which is now popular in the application of image reconstruction [9] and data analysis [10]. For the application in dose painting, proper regularization terms need to be integrated during the optimization of dose efficiency.

Radiation response is strongly influenced by tumor oxygenation, which is usually assessed using [^{18}F]FMISO PET images. Kinetic modeling is often used to assess tracer pharmacokinetic parameters [11] and has been reported to be of advantage compared to direct evaluation of static [^{18}F]FMISO PET images [11]. However, it is usually not possible to quantitatively relate the estimated kinetic parameters directly to the underlying oxygen tensions due to the complex tracer transport and metabolism process in the heterogeneous tumor microenvironment. On the other hand, a recently developed simulation model based on reaction-diffusion equations by Kelly et al. [12] is able to simulate the distributions of tracer and oxygen under the same tumor microenvironment, thus it is possible to explore the relation between [^{18}F]FMISO imaging and oxygen tension. An extension of the Kelly model called flow-limited oxygen-dependent (FLOD) model has been validated with preclinical PET data [13].

This paper proposes a framework to optimize the dose distribution based on quantitative analysis of [^{18}F]FMISO PET image series. We developed a sparse dose painting (SDP) algorithm by optimizing a linear combination of two terms corresponding to the efficiency and total variation of the dose distribution under the constraint of constant mean dose, which comprises the therapy outcome and the treatment complexity. The radiosensitivity parameter, oxygen tension, is estimated using one dual-pass kinetic-oxygen mapping strategy, which is achieved by simulating the transport and metabolism of oxygen and tracer under the same condition and setting up a kinetic-oxygen mapping curve based on the simulation results of various tumor microenvironments.

2 Sparse Dose Painting

One aim in dose painting is to compute a proper dose distribution $\mathbf{D}$ in a tumor domain Ω based on the estimated oxygen tensions from [^{18}F]FMISO PET images. Given a certain prescribed mean dose D_{mean}, treatment planning attempts to optimize the dose $D(\mathbf{u})$ of each image voxel $\mathbf{u}, \mathbf{u} \in \Omega$ and achieve a maximum treatment efficiency, i.e. a minimum fraction of surviving cells after irradiation (surviving fraction SF). However, IMRT is usually done with a limited number of beams and restricted range of intensity modulation leading to a sparse dose prescription. An additional term $\|\mathbf{D}\|_{TV} = \sum_{i,j,k\in\Omega}((\nabla_1 D_{ijk})^2+(\nabla_2 D_{ijk})^2+(\nabla_3 D_{ijk})^2)$ is introduced to regularize the total variation (TV) and to reduce the complexity of the dose distribution. Here ∇_1,

∇_2 and ∇_3 denote the forward finite difference operators on the three coordinates, respectively. The optimization metric can be described in the following mathematic form,

$$\arg\min_{\mathbf{D}} \{log(\|SF(\mathbf{D})\|_1) + \lambda \frac{\|\mathbf{D}\|_{TV}}{N_{voxel}}\} \ \ subject\ to \ \ \text{mean}(\mathbf{D}) = D_{mean} \quad (1)$$

where λ is the regularization parameter, which is patient-specific concerning the patient tolerance, tumor geometry and the topology with the surrounding normal tissue.

The surviving fraction $SF(\mathbf{u})$ of each voxel is determined by the linear-quadratic model, which is an empirical formula for the predication of radiation response [14]:

$$\begin{aligned} SF(\mathbf{u}) &= e^{-\alpha(\bar{p}(\mathbf{u}))D(\mathbf{u}) - \beta(\bar{p}(\mathbf{u}))D(\mathbf{u})^2} \\ \alpha(\bar{p}(\mathbf{u})) &= \alpha_0 \frac{m\bar{p}(\mathbf{u}) + K}{\bar{p}(\mathbf{u}) + K} \\ \beta(\bar{p}(\mathbf{u})) &= \beta_0 (\frac{m\bar{p}(\mathbf{u}) + K}{\bar{p}(\mathbf{u}) + K})^2 \end{aligned} \quad (2)$$

α and β are the radiation sensitivity coefficients, which are determined by the mean oxygen tension $\bar{p}$ of imaging voxels $\mathbf{u}$. α_0, β_0, m and K are constants. Eqn 1 is solved using a conjugate gradient method [15] with the initialization of DPBN applied to the results of kinetic modeling.

3 Dual-Pass Kinetic-Oxygen Mapping

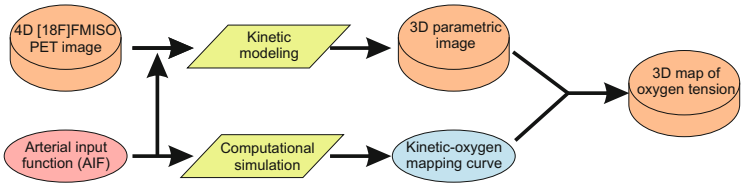


Fig. 1. The procedure of dual-pass kinetic-oxygen mapping.

One prerequisite to solve Eqn. 1 is to quantitatively estimate the radiosensitivity parameter, oxygen tension, from the acquired [^{18}F]FMISO PET images. Kinetic modeling is often used for quantitative analysis of PET images [11,16]. However, the estimated parameters are still not directly related to oxygen tension. On the other hand, computational simulation of [^{18}F]FMISO PET imaging has advantages in revealing the correlation between [^{18}F]FMISO kinetics and oxygen tension by computing their distribution under the same tumor microenvironment [12]. We developed a strategy to combine both the forward and backward modeling and to quantify the oxygen tension from dynamic PET images. The procedure is sketched in Fig. 1. A kinetic-oxygen mapping curve is derived by fitting the computational simulation results using the following formula,

$$k_3 = \frac{a}{\bar{p} + b} + c \quad (3)$$

where k_3 is the estimated kinetic parameter using a typical irreversible two-compartment model [16] and $\bar{p}$ is the mean oxygen tension of a voxel $\mathbf{u}$. Constants of a, b, c reflect the integral of the heterogeneous tracer binding procedure. They vary for different patients or animals due to different metabolic properties. The mapping curve of Eqn. 3 is then applied to transform the parametric image to the oxygen tension image.

Reaction-Diffusion Simulation Model

To identify the constants of Eqn. 3, we use an extension of the Kelly model [12] called flow-limited oxygen-dependent (FLOD) model to simulate the relation between the oxygen consumption and the [^{18}F]FMISO binding inside an imaging voxel in heterogeneous tumor microenvironments [13]. Tumor microenvironments are modeled as 2D grids (grid size 10 μm) spreading with heterogeneously distributed vessels. The transport and metabolism of oxygen and [^{18}F]FMISO molecules are modeled as a diffusion-reaction procedure with an assumption of flow-limited extraction. Given the physical and metabolic properties of oxygen and [^{18}F]FMISO molecules measured from *in vitro* experiments, the distribution of oxygen and [^{18}F]FMISO in tumor areas can be computed by solving the following reaction-diffusion equations:

$$\begin{aligned}
\frac{\partial Tr_{free}(\mathbf{x},t)}{\partial t} &= D_{Tr}\nabla^2 Tr_{free}(\mathbf{x},t) - k_{binding}(\mathbf{x})Tr_{free}(\mathbf{x},t) \\
\frac{\partial Tr_{bound}(\mathbf{x},t)}{\partial t} &= k_{binding}(\mathbf{x})Tr_{free}(\mathbf{x},t) \\
k_{binding}(\mathbf{x}) &= \frac{k_a}{p(\mathbf{x}) + k_b} \\
D_{O_2}\nabla^2 p(\mathbf{x}) &= -q_{max}\frac{p(\mathbf{x})}{p(\mathbf{x}) + k_p}
\end{aligned} \tag{4}$$

where $Tr_{free}(\mathbf{x},t)$ describes the concentration of the free ligand for a spatial location $\mathbf{x}$ at time t, and $Tr_{bound}(\mathbf{x},t)$ is the concentration of the bound ligand. $k_{binding}(\mathbf{x})$ is the binding rate of the tracer, which relates to the cellular oxygen tension $p(\mathbf{x})$. D_{Tr} and D_{O_2} are diffusion coefficients of [^{18}F]FMISO and oxygen. q_{max} and k_p are oxygen metabolic constants. k_a and k_b are constants determined by [^{18}F]FMISO pharmacodynamics. Note that these cellular parameters are different from the macroscopic parameters a, b, c in Eqn. 3. Further, the microscopic binding parameter $k_{binding}$ is different from the macroscopic estimation parameter k_3 using kinetic modeling.

4 Experiments and Results

We tested our algorithm on one data set of a patient with head and neck cancer, which was scanned with CT and a dynamic [^{18}F]FMISO PET. The image series consist of 30 continuous frames of the first 15 min post injection (p. i.) of [^{18}F]FMISO and an additional frame at 2 hours p. i.. Two blood samples were acquired after the PET scan. The arterial input function (AIF) is derived using an image based method by selection of a carotid artery in the [^{18}F]FMISO PET images followed by a correction of the partial

volume effect. The measured blood activity is applied to correct the spill over effect by replacing the tail part of the image derived AIF. A three exponential model is applied to smooth the combined AIF [17]. The PET images were registered to the CT images using a mutual information algorithm. The tumor regions (primary tumor and affected lymph nodes) were outlined manually by an experienced radiation oncologist. Fig. 2a shows a slice of the fused CT and PET images.

Fig. 2 shows the results of the analysis of the dynamic [^{18}F]FMISO PET images using the dual-pass kinetic-oxygen mapping method. Fig. 2b depicts the estimated k_3 image of the slice in Fig. 2a. Around 400 PET imaging voxels (voxel size: 4×4 mm^2) were simulated based on the extracted AIF. The k_3 values of the simulated TACs and the corresponding mean oxygen tensions are plotted in Fig. 2c. The kinetic-oxygen mapping curve is generated by fitting the simulation results using Eqn. 3. A slice of the map of the oxygen tension after kinetic-oxygen mapping is illustrated in Fig. 2d.

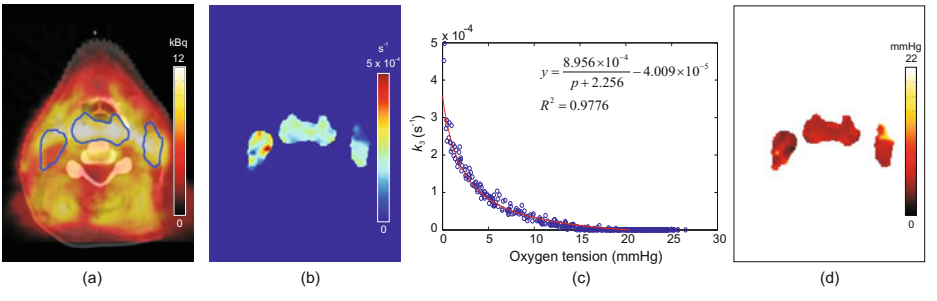


Fig. 2. Results of the dual-pass kinetic-oxygen mapping: (a) slice of the fusion of the CT image and the [^{18}F]FMISO PET image 2 hours p.i.; (b) parametric image of k_3 using pixel-wise irreversible two-compartment model within the tumor regions for slice a; (c) plot of the values of the kinetic parameters k_3 and the mean oxygen tensions of 400 simulated voxels and the resulting kinetic-oxygen mapping curve fitted using Eqn. 3; (d) reconstructed oxygen tension map within the tumor regions for slice a.

SDP with different λ values was tested and compared with direct DPBN, which uses a linear transformation of static [^{18}F]FMISO images to the dose prescription [6][1]. The prescribed dose distributions using SDP and DPBN were compared for the prescribed mean dose of $D_{mean} = 2.3$ Gy for a single fraction within the tumor regions here. One example slice of different resulting dose prescriptions is displayed in Fig.3a-e. Even without sparsity regularization, dose painting based on the dual-pass kinetic-oxygen quantification (Fig. 3a) generates significantly different results compared to DPBN (Fig. 3e). It confirms that a direct evaluation of PET image may not be sufficient to assess the underlying oxygen tension [11].

Fig. 3a-d show the comparison of increasing λ values, which will reduce the complexity of dose prescription correspondingly. The regulated dose prescription has several advantages: (1) it will diminish the burden of inverse planning; (2) it will reduce

[1] The DPBN presented in [7] is not considered in this paper. The comparison of different tumor control probability models is not feasible in this framework.

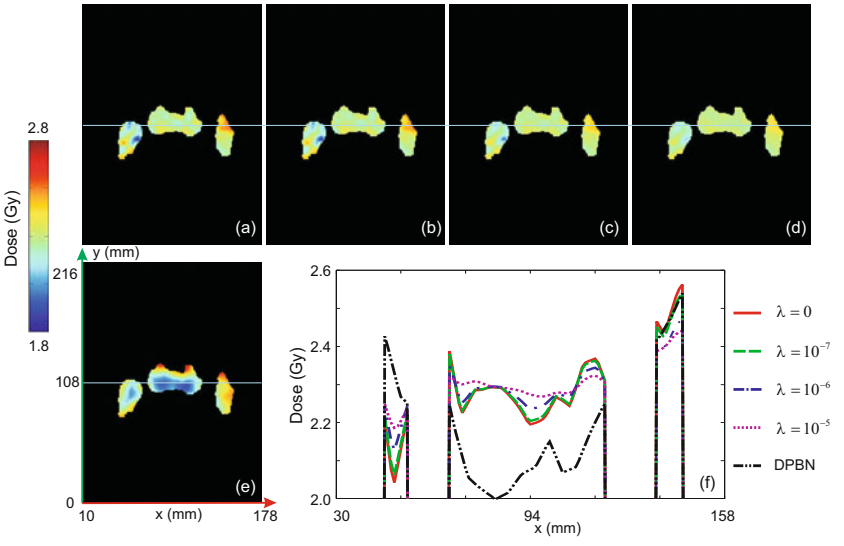


Fig. 3. Comparison of SDP with different λ and DPBN for the patient case: (a-d) One slice of SDP dose prescription, $\lambda = 0, 10^{-7}, 10^{-6}, 10^{-5}$, respectively; (e) One slice of DPBN dose prescription; (f) A plot of the dose values along the blue lines in the middle of the selected slice shown in a-e.

Table 1. Comparison of the resulting average surviving fraction of different methods

DPBN	SDP			
	$\lambda = 0$	$\lambda = 10^{-7}$	$\lambda = 10^{-6}$	$\lambda = 10^{-5}$
2.99×10^{-4}	2.61×10^{-6}	4.69×10^{-6}	1.17×10^{-4}	1.22×10^{-4}

λ: regularization parameter; SDP: sparse dose painting; DPBN: dose painting by number.

the number of beams and the complexity of fluence modulation leading to a reduced treatment time; (3) it will improve the quality of normal tissues sparing; besides treating the tumor, radiotherapy needs to avoid too much side effects in normal tissues. For a constant total dose, SDP reduces extremely high dose prescriptions in the tumor (shown in Fig. 3). Thus it is also possible to reduce the probability of hot spots in the normal tissue. Overall, the sparsity regularization will increase the feasibility of dose painting. The corresponding dose efficiency after 30 fractions is listed in Tab 1. SDP has a better dose efficiency compared to DPBN due to the improved quantification of the oxygen tensions. Note, due to the large values of the TV norm, the λ values are quite small here to ensure a proper regularization. A suitable λ value is determined by various factors, such as the volume and geometry of the tumor and the topology with the surrounding normal tissues. It can be adapted in real applications and will be optimized based on statistical analysis of further studies.

5 Conclusion

This paper develops a novel framework to optimize the dose prescription in dose painting based on quantitative analysis of [^{18}F]FMISO PET images. In particular, we made the following two contributions: (1) By incorporating a realistic PET imaging simulation model, this paper sets up a dual-pass kinetic-oxygen mapping to transform the estimated kinetic parameters to physiological parameters of the tumor microenvironment. (2) A sparse dose painting algorithm is developed to optimize the dose efficiency as well as the dose modulation. This is achieved by adding a sparsity regularization term during the optimization of dose efficiency with the constraint of constant mean dose.

Our method is limited by the indirect quantification by integrating computational simulation into tracer kinetic modeling, which might induce bias and errors. (1) Although the FLOD simulation model is in good agreement with preclinical experimental data, for clinical applications it is necessary to further adjust and validate the model settings. (2) The two-pass quantification procedure may accumulate and amplify errors in some situations. Direct matching of reaction-diffusion simulation to imaging data is desired [2]. However, the tracer uptake procedure is determined by both the spatial location and the perfusion of vessels spreading into tissues discretely and randomly, which is generally not differentiable. Nevertheless, our method offers the potential to improve the treatment outcome of dose painting. Future work will focus on the further optimization of dose prescription in the clinical target volume by integrating tumor growth parameters [2] in a tumor control probability model and on the testing and integration of the algorithm in real therapy planning systems.

Acknowledgements. Supported by BMBF MobiTUM project (01EZ0826) and DFG Cluster of Excellence: Munich-Centre for Advanced Photonics.

References

1. Konukoğlu, E., Clatz, O., Bondiau, P.-Y., Delingette, H., Ayache, N.: Extrapolating tumor invasion margins for physiologically determined radiotherapy regions. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4190, pp. 338–346. Springer, Heidelberg (2006)
2. Konukoglu, E., Clatz, O., Menze, B.H., Stieltjes, B., Weber, M.A., Mandonnet, E., Delingette, H., Ayache, N.: Image Guided Personalization of Reaction-Diffusion Type Tumor Growth Models Using Modified Anisotropic Eikonal Equations. IEEE Trans. Med. Imaging 29, 77–95 (2010)
3. Lu, C., Chelikani, S., Chen, Z., Papademetris, X., Staib, L.H., Duncan, J.S.: Integrated segmentation and nonrigid registration for application in prostate image-guided radiotherapy. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6361, pp. 53–60. Springer, Heidelberg (2010)
4. Ling, C., Humm, J., Larson, S., Amols, H., Fuks, Z., Leibel, S., Koutcher, J.: Towards Multidimensional Radiotherapy (MD-CRT): Biological Imaging and Biological Conformality. Int. J. Radiat. Oncol. Biol. Phys. 47, 551–560 (2000)
5. Simari, P.D., Wu, B., Jacques, R., King, A., McNutt, T., Taylor, R., Kazhdan, M.M.: A statistical approach for achievable dose querying in IMRT planning. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6363, pp. 521–528. Springer, Heidelberg (2010)

6. Alber, M., Paulsen, F., Eschmann, S.M., Machulla, H.J.: On Biologically Conformal Boost Dose Optimization. Phys. Med. Biol. 35, N31–N35 (2003)
7. Thorwarth, D., Eschmann, S., Paulsen, F., Alber, M.: Hypoxia Dose Painting by Numbers: a Planning Study. Int. J. Radiat. Oncol. Biol. Phys. 68, 291–300 (2007)
8. Vaupel, P.: Tumor Microenvironmental Physiology and its Implications for Radiation Oncology. Semin. Radiat. Oncol. 14, 198–206 (2004)
9. Huang, J., Zhang, S., Metaxas, D.N.: Efficient MR image reconstruction for compressed MR imaging. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6361, pp. 135–142. Springer, Heidelberg (2010)
10. Avants, B.B., Cook, P.A., McMillan, C., Grossman, M., Tustison, N.J., Zheng, Y., Gee, J.C.: Sparse unbiased analysis of anatomical variance in longitudinal imaging. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6361, pp. 324–331. Springer, Heidelberg (2010)
11. Thorwarth, D., Eschmann, S., Paulsen, F., Alber, M.: A Kinetic Model for Dynamic ^{18}F-Fmiso PET Data to Analyse Tumour Hypoxia. Phys. Med. Biol. 50, 2209–2224 (2005)
12. Kelly, C., Brady, M.: A Model to Simulate Tumour Oxygenation and Dynamic [^{18}F]-Fmiso PET Data. Phys. Med. Biol. 51, 5859–5873 (2006)
13. Shi, K., Bayer, C., Maftei, C.A., Gaertner, F.C., Astner, S.T., Wilkens, J.J., Nüsslin, F., Vaupel, P., Ziegler, S.I.: A Flow-limited Oxygen-dependent Diffusion Model with Heterogeneous Perfusion for Quantitative Analysis of Dynamic [^{18}F]Fmiso PET Uptake in Tumors (submitted)
14. Wenzl, T., Wilkens, J.J.: Modelling of the Oxygen Enhancement Ratio for Ion Beam Radiation Therapy. Phys. Med. Biol. 56, 3251–3268 (2011)
15. Nocedal, J., Wright, S.J.: Numerical Optimization, 2nd edn. Springer, Heidelberg (2000)
16. Shi, K., Souvatzoglou, M., Astner, S.T., Vaupel, P., Nüsslin, F., Wilkens, J.J., Ziegler, S.I.: Quantitative Assessment of Hypoxia Kinetic Models by a Cross-Study of Dynamic ^{18}F–FAZA and ^{18}O-H_2O in Patients with Head and Neck Tumors. J. Nucl. Med. 51, 1386–1394 (2010)
17. Feng, D., Huang, S., Wang, X.: Models for Computer Simulation Studies of Input Functions for Tracer Kinetic Modeling with Positron Emission Tomography. Int. J. Biomed. Comput. 32, 95–110 (1993)

Robust Estimation of Kinetic Parameters in Dynamic PET Imaging

Fei Gao[1], Huafeng Liu[2,1], and Pengcheng Shi[1]

[1] Golisano College of Computing and Information Sciences, Rochester Institute of Technology, Rochester, NY, 14623, USA
[2] State Key Laboratory of Modern Optical Instrumentation, Zhejiang University, Hangzhou, 310027, China

Abstract. Dynamic PET imaging provides important information for biological research, clinical diagnosis and pharmacokinetic analysis through kinetic modeling and data-driven parameter estimation. Kinetic parameters quantitatively describe dynamic material exchange and metabolism of radiotracers in plasma and tissues. While many efforts have been devoted to estimate kinetic parameters from dynamic PET, the poor statistical properties of the measurement data in low count dynamic acquisition and the uncertainties in estimating the arterial input function have limited the accuracy and reliability of the kinetic parameter estimation. Additionally, the quantitative analysis of individual kinetic parameters is not yet implemented. In this paper, we present a robust kinetic parameter estimation framework which is robust to both the poor statistical properties of measurement data in dynamic PET and the uncertainties in estimated arterial input function, and is able to analyze every single kinetic parameter quantitatively. The strategy is optimized with robust H_∞ estimation under minimax criterion. Experiments are conducted on Monte Carlo simulated data set for quantitative analysis and validation, and on real patient scans for assessment of clinical potential.

1 Introduction

Dynamic Positron Emission Tomography (PET) is a molecular imaging technique used to monitor the spatiotemporal distribution of a radiotracer *in vivo* and provides important information for biological research, clinical diagnosis and pharmacokinetic analysis [1,2]. PET offers good promise for accelerating the process from pre-clinical discover to Phase III studies of drug development, which benefits from various radiolabelled biochemicals and well established metabolism models. The accurate quantitative descriptions (kinetic parameters) of the interaction of the drug with a desired binding site and the neurotransmitter concentration will significantly advance the studies of drug discovery and development.

The typical approach of estimating kinetic parameters is first to reconstruct activity distributions, and then to fit the calculated time activity curve (TAC) to a predefined kinetic model. However, the practicability of this kind of approaches relies on the accuracy of image reconstruction. There are also many

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 492–499, 2011.

efforts trying to estimate from projection data directly by introducing either linear or nonlinear models [3], however, the optimization algorithms are generally very complicated to implement. Kamasak *et al.* applied coordinate descent optimization but which is still limited in specific models [4]. Wang *et al.* applied an generalized algorithm for reconstruction of parametric images [5]. However, in dynamic PET imaging, the low count data acquisition procedures make the statistical properties of measurement data very poor, moreover, the uncertainties introduced by various data corrections make the noise properties complicated, and will lead to a suboptimal estimation [6]. Additionally, difficulties in arterial sampling and uncertainties during input function fitting will result in errors in kinetic parameter estimation. Furthermore, the quantitative estimation and analysis of the dynamic change of individual kinetic parameter, which is not yet implemented, will definitely lead to better understandings of tissue biology [7].

In this paper, we concentrate on the development of a robust kinetic parameter estimation framework from projection data directly. PET measurement equation is successfully transformed and combined with kinetic models, which makes it very flexible to satisfy different kinetic models. The framework is robust to both poor statistical properties of measurement data and uncertainties in arterial input function estimation when no blood sampling exists, and is able to analyze every single kinetic parameter quantitatively. The strategy is optimized by robust H_∞ filter under minimax criterion. Experiments are conducted with Monte Carlo simulated dynamic data for statistical analysis and validation, and on real patient scans for assessment of clinical potential.

2 Method

2.1 Modeling of Tracer Kinetics

In this paper, a general two-tissue three-compartment model is adopted to describe regional tracer kinetics as shown in Fig.1, where C_P(pmol/ml) is arterial concentration of radiotracer, C_F and C_B (pmol/ml) are the concentrations of non–specific binding and specific binding tracers in tissues. Parameters k_1, k_2, k_3 and $k_4(min^{-1})$ specify radiotracer transport rates. The time variation of kinetic model in voxel i can be denoted by first–order differential equations as:

$$\frac{dC_{Fi}(t)}{dt} = k_{1i}(t)C_{Pi}(t) + k_{4i}(t)C_{Bi}(t) - (k_{2i}(t) + k_{3i}(t))C_{Fi}(t) \tag{1}$$

$$\frac{dC_{Bi}(t)}{dt} = k_{3i}(t)C_{Fi}(t) - k_{4i}(t)C_{Bi}(t) \tag{2}$$

2.2 Modeling of Dynamic PET Measurement with Tracer Kinetics

Dynamic PET imaging involves a sequence of contiguous acquisition with different temporal resolutions, which can be formulated as a projection transform:

$$y(t) = Dx(t) + e(t) \tag{3}$$

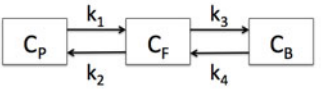


Fig. 1. Two–tissue three–compartment model

Here, $y(t)$ is the projection data and $x(t) = \{x_i(t)|i = 1, \cdots, n\}^T$ is the activity concentration at time frame t. n is the total number of voxels. D is the system probability matrix. $e(t)$ is the overall measurement uncertainties. Here we will transform Eqn.(3) to accommodate kinetic models. Firstly, activity concentration x will be the combination of C_F and C_B, then Eqn.(3) will be

$$y(t) = \begin{bmatrix} D & D \end{bmatrix} \begin{bmatrix} C_F(t) \\ C_B(t) \end{bmatrix} + e(t) \tag{4}$$

where $C_F(t) = \{C_{Fi}(t)|i = 1, \cdots, n\}^T$ and $C_B(t) = [C_{Bi}(t)|i = 1, \cdots, n\}^T$. After the dynamic change of measurement $\frac{dy_i(t)}{dt}$ being deduced, we substitute the differential equations Eqn.(1) and (2) and do a simple transformation to arrange 4 kinetic parameters $(k1, k2, k3, k4)$ in a column vector will yield

$$\frac{dy_i(t)}{dt} = \begin{bmatrix} D & D \end{bmatrix} \begin{bmatrix} \vdots \\ \frac{dC_{Fi}(t)}{dt} \\ \vdots \\ \frac{dC_{Bi}(t)}{dt} \\ \vdots \end{bmatrix} + e_i^{'}(t) = \begin{bmatrix} D & D \end{bmatrix} \begin{bmatrix} & \vdots & & \\ C_{Pi}(t) & -C_{Fi}(t) & -C_{Fi}(t) & C_{Bi}(t) \\ & \vdots & & \\ 0 & 0 & C_{Fi}(t) & -C_{Bi}(t) \\ & \vdots & & \end{bmatrix} \begin{bmatrix} k_{1i}(t) \\ k_{2i}(t) \\ k_{3i}(t) \\ k_{4i}(t) \end{bmatrix} + e_i^{'}(t) \tag{5}$$

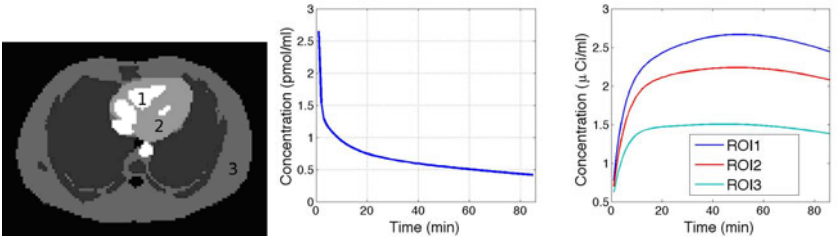
By denoting $R_i(t) = \begin{bmatrix} & \vdots & & \\ C_{Pi}(t) & -C_{Fi}(t) & -C_{Fi}(t) & C_{Bi}(t) \\ & \vdots & & \\ 0 & 0 & C_{Fi}(t) & -C_{Bi}(t) \\ & \vdots & & \end{bmatrix}$ and $S_i(t) = \begin{bmatrix} k_{1i}(t) \\ k_{2i}(t) \\ k_{3i}(t) \\ k_{4i}(t) \end{bmatrix}$,

we can get the dynamic change of total measurement data from all voxels as

$$\frac{dy(t)}{dt} = \sum_{i=1}^{n} \frac{dy_i(t)}{dt} = \begin{bmatrix} D & D \end{bmatrix} \begin{bmatrix} R_1(t) \cdots & R_i(t) \cdots & R_n(t) \end{bmatrix} \begin{bmatrix} S_1(t) \\ \vdots \\ S_i(t) \\ \vdots \\ S_n(t) \end{bmatrix} + e^{'}(t) \tag{6}$$

$$= \begin{bmatrix} D & D \end{bmatrix} R(t)S(t) + e^{'}(t)$$

Now we have set up the relationship between the change of measurement data and kinetic parameters directly by Eqn.(6).



(a) (b) (c)

Fig. 2. (a) Zubal phantom; (b) Blood input function; (c) Output TACs of 3 ROIs

2.3 Robust Solution by H_∞ Optimization Under Minimax Criterion

Gao et al.[8] present an efficient and robust estimation framework for activity reconstruction of dynamic PET imaging. No statistical assumptions needed make it robust to the poor statistical properties in low count acquisition and system noises. We will transform our problem to fit that solution framework. Since kinetic parameters are generally assumed to be constant, we can set:

$$S(t+1) = S(t) + v(t) \tag{7}$$

Here, v is possible disturbances. With Eqn.(6) and Eqn.(7), the corresponding minimax performance equation will be "$\min_{S\in L} \max_{e\in E, v\in V} F(S, e, v)$", where L, E and V are the sets of solutions, uncertainties of measurement and state transition. As an iterative solution, we also define a linear combination of $S(t)$ as "$z(t) = g(S(m), v(m))$ where $m = 1, 2...t$", then objective function J will be

$$J = \frac{\sum \|z(t) - \hat{z}(t)\|^2_{Q(t)}}{\|S(0) - \hat{S}(0)\|^2_{p_o^{-1}} + \sum \left(\|v(t)\|^2_{V(t)^{-1}} + \|e(t)\|^2_{E(t)^{-1}}\right)} \tag{8}$$

where the notation $\|x\|^2_G$ is defined as the square of the weighted (by G) L_2 norm of x. p_o, $E(t)$, $V(t)$ and $Q(t)$ are weighting matrices. $\hat{S}(0)$ is the initialization of x. More detailed settings and initialization of parameters can be found in [9].

3 Experiments

3.1 Monte Carlo Simulated Dynamic PET Data

The first dynamic PET data set is from Monte Carlo simulation. The simulated PET scanner is Hamamatsu SHR74000. The phantom used here is Zubal thorax phantom. One sample slice is shown in Fig. 2(a). 3 regions including heart, muscle and chest wall are selected as ROI1–3. The experiment is a dynamic ^{18}F–FDG study with the compartment model in Sec 2.1 for imaging glucose metabolism. The TACs of 3 ROIs are generated by analytical Feng's input function:

$$C_P^{FDG}(t) = (A_1 t - A_2 - A_3)e^{-\lambda_1 t} + A_2 e^{-\lambda_2 t} + A_3 e^{-\lambda_3 t} \tag{9}$$

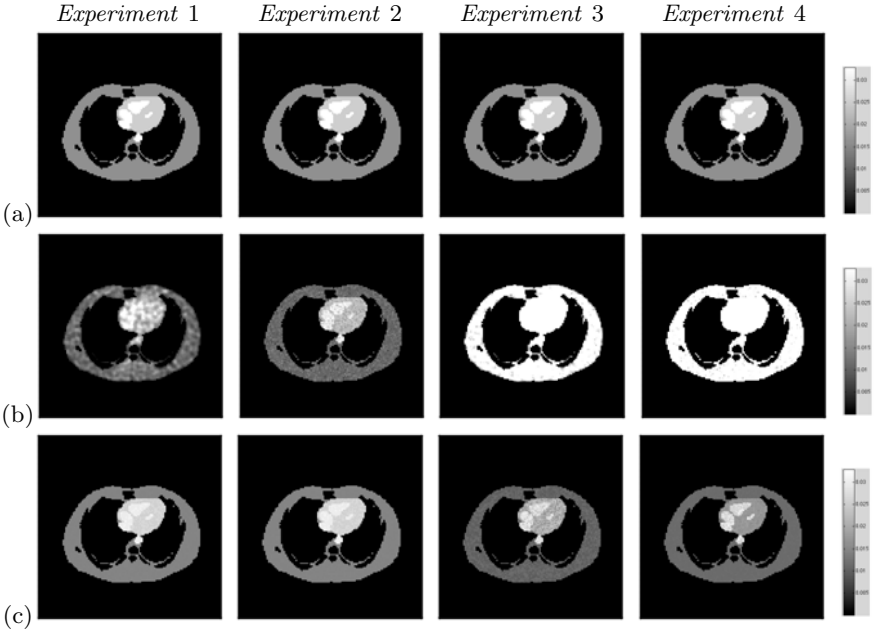


Fig. 3. Influx rate maps. (a) Ground Truth; (b) WLSCF; (c) Our Method.

The parameters λ_i and A_i used here are $A_1 = 85.112\mu Ci/mL/min$, $A_2 = 2.081\mu Ci/mL$, $A_3 = 2.188\mu Ci/mL$, $\lambda_1 = 4.1339min^{-1}$, $\lambda_2 = 0.01043min^{-1}$ and $\lambda_3 = 0.1191min^{-1}$. The dynamic acquisition consists of 85 frames: $15\times0.2min$, $20\times0.5min$, $40\times1min$ and $10\times3min$. Kinetic parameters in simulations are, ROI1: 0.102, 0.130, 0.062, 0.0068; ROI2: 0.082, 0.102, 0.045, 0.004; ROI3: 0.064, 0124, 0.042, 0.0035. Calculated input function and TACs are shown in Fig.2 (b) and (c). The differences of total counts between time frames are around 1k.

4 experiments (*Experiment* 1-4) are conducted under different conditions. Since input function is required and no blood sampling is desired, firstly, 4 experiments are divided into 2 groups: the first group (*Experiment* 1 and 2) uses a perfect input function, while the second group (*Experiment* 3 and 4) uses an imperfect input function generated by image-derived sampling method. Secondly, due to poor statistical properties of measurement data in low count dynamic PET imaging, spatial constrains are also introduced. In *Experiment* 1 and 3, every voxel is estimated separately. For a comparison, *Experiment* 2 and 4 are performed with "local average" constraints to improve statistical properties, which means the measurement data in one voxel is averaged by surrounding voxels. Besides our method, a typical Weighted Least Square Curve Fit (WLSCF) method provided by COMKAT toolbox, is conducted as comparison. To evaluate the experiments quantitatively, statistical analysis on estimated results against true values is performed. Let N_p be the total number of voxels or ROIs, SE_i and ST_i be the estimation of ith kinetic parameter and corresponding true value, then define $Bias = \frac{1}{N_p}\sum_{i=1}^{N_p}\frac{(SE_i - ST_i)}{ST_i}$ and $Std = \sqrt{\frac{1}{N_p-1}\sum_{i=1}^{N_p}(\frac{SE_i-ST_i}{ST_i})^2}$. We also

Table 1. The calculated Bias and Std of *Experiment* 1 (top) and 2 (bottom)

	Bias±Std	k1	k2	k3	k4
ROI1	WLSCF	0.10706 ± 0.21072	0.17943 ± 0.30803	-0.09460 ± 0.25205	-0.26725 ± 0.40846
	Our Method	-0.04794 ± 0.05450	-0.16225 ± 0.17368	-0.20677 ± 0.22111	-0.14092 ± 0.15652
ROI2	WLSCF	-0.01347 ± 0.22397	0.24409 ± 0.32029	-0.03134 ± 0.22162	0.16834 ± 0.65988
	Our Method	-0.00459 ± 0.04190	0.10929 ± 0.11838	0.18639 ± 0.19926	0.09558 ± 0.14263
ROI3	WLSCF	-0.16228 ± 0.29534	0.37212 ± 0.47103	0.07289 ± 0.24139	0.29611 ± 0.81515
	Our Method	-0.12107 ± 0.14058	-0.02826 ± 0.08700	0.00574 ± 0.00614	0.19748 ± 0.21488

	Bias±Std	k1	k2	k3	k4
ROI1	WLSCF	0.00502 ± 0.08444	0.09966 ± 0.12406	-0.08475 ± 0.11367	-0.11420 ± 0.20296
	Our Method	-0.01904 ± 0.02867	-0.03211 ± 0.03446	-0.07005 ± 0.07490	0.10215 ± 0.10967
ROI2	WLSCF	-0.01809 ± 0.08325	0.21458 ± 0.23228	0.03688 ± 0.09033	0.13241 ± 0.32270
	Our Method	-0.06743 ± 0.07298	-0.03760 ± 0.04064	-0.00786 ± 0.00840	0.05390 ± 0.05763
ROI3	WLSCF	-0.16799 ± 0.20050	0.33093 ± 0.34760	0.05029 ± 0.09786	0.28117 ± 0.42891
	Our Method	-0.05638 ± 0.06740	0.08000 ± 0.09407	0.04364 ± 0.04665	-0.14543 ± 0.15650

calculate the influx rate map (Rate) which is related to the glucose metabolic rate as $Rate = \frac{k_1 k_3}{k_2 + k_3}$. Parametric maps also show visual comparisons.

Table 1 and 2 show calculated Bias and Std of 4 experiments in 2 groups. Fig.3 shows the calculated influx rate maps. Table 1 indicates that WLSCF can achieve acceptable results with perfect input function, and the local average constraint introduced in *Experiment* 2 improves Std of all estimated parameters by improving the statistical properties of measurement data. The Bias and Std of results from our methods are better than that from WLSCF. The results also show the robustness of our method in dealing with the poor statistical properties of measurement data. The statistical analysis in Table 2 shows that with input function from estimation, WLSCF leads to worse results indicated by increased Std values in both experiments, especially k_1 and k_4. k_1 is the rate constant indicating the radiotracer exchange from arterial to tissue, which is highly affected by the accuracy of estimated input function, and k_4 is the FDG dephosphorylation rate and very sensitive to the system. The influx rate maps also show obvious overestimation of WLSCF. However, our method still achieves robust estimation results of all parameters in both experiments and yields acceptable Std with the imperfect input function. And the calculated Std shows that some parameter estimations by our method with imperfect input function are even comparable with that by WLSCF with perfect input function.

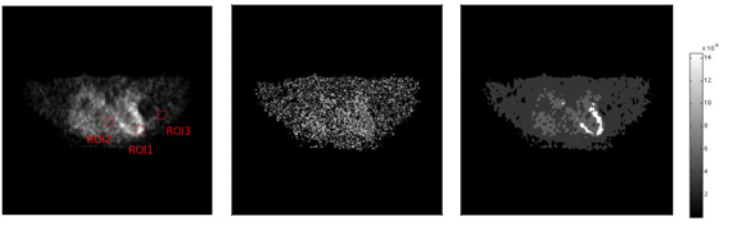
3.2 Experiments with Data from Real Patient Scan

The real patient data in this study is a dynamic PET scan acquired from a 28-year-old, 75kg male volunteer. The scanner used is Hamamatsu SHR-22000 whole body PET scanner. 10 mCi ^{18}F-FDG is injected and a dynamic acquisition of the thoracic cavity starts just after injection. The acquisition consists of 40 time frames: $20\times0.5min$, $15\times1min$, and $5\times2min$. The input function is estimated by the image-derived method. Since the influx rate maps have direct relationships

Table 2. The calculated Bias and Std of *Experiment* 3 (top) and 4 (bottom)

	Bias±Std	k1	k2	k3	k4
ROI1	WLSCF	1.05780 ± 1.14375	-0.09464 ± 0.30377	0.37156 ± 0.68273	0.12470 ± 0.94807
	Our Method	-0.27954 ± 0.30441	-0.15235 ± 0.16289	0.05914 ± 0.06328	-0.12877 ± 0.13901
ROI2	WLSCF	1.25494 ± 1.31110	0.11049 ± 0.38121	0.45580 ± 0.67132	0.24023 ± 1.13308
	Our Method	-0.36468 ± 0.39731	-0.16074 ± 0.17187	0.21405 ± 0.22883	0.30306 ± 0.32609
ROI3	WLSCF	1.20747 ± 1.28191	-0.12640 ± 0.26492	0.00213 ± 0.28536	0.48420 ± 1.51799
	Our Method	-0.39567 ± 0.42691	-0.13915 ± 0.25402	-0.05089 ± 0.05440	0.24891 ± 0.27461

	Bias±Std	k1	k2	k3	k4
ROI1	WLSCF	1.14242 ± 1.20911	0.17117 ± 0.33201	0.66864 ± 0.82730	-0.17832 ± 0.51223
	Our Method	-0.19232 ± 0.21434	0.17155 ± 0.18339	0.22930 ± 0.24531	-0.06705 ± 0.07176
ROI2	WLSCF	1.24255 ± 1.27450	0.15610 ± 0.28138	0.50822 ± 0.61257	-0.00525 ± 0.78607
	Our Method	-0.26658 ± 0.28651	0.25757 ± 0.27536	0.12868 ± 0.13816	-0.26835 ± 0.28689
ROI3	WLSCF	1.25741 ± 1.27605	-0.13817 ± 0.22765	-0.00537 ± 0.11263	0.19770 ± 1.04567
	Our Method	-0.22002 ± 0.23667	-0.22338 ± 0.24597	-0.00636 ± 0.00680	0.16175 ± 0.17297



(a) (b) WLSCF (c) Our Method

Fig. 4. (a) Reference activity map with 3 ROIs; (b), (c) Influx rate map

Table 3. Estimated kinetic parameters (top) and their MSE (bottom)

ROI1				ROI2				ROI3			
k1	k2	k3	k4	k1	k2	k3	k4	k1	k2	k3	k4
0.0805	0.4659	0.1020	0.0058	0.0493	0.9758	0.1198	0.0059	0.0094	0.1984	0.0973	0.0079

ROI1				ROI2				ROI3			
k1	k2	k3	k4	k1	k2	k3	k4	k1	k2	k3	k4
0.1426	0.2030	0.1326	0.2018	0.2770	0.1438	0.1445	0.2047	0.3284	0.2236	0.1873	0.1989

with activity distribution, so we choose an activity map from static scan as reference to extract ROIs. Fig. 4(a) shows the reference activity map and 3 ROIs are selected including left ventricular, heart muscle and body surface. Table 3 shows the estimated kinetic parameters by our method and their mean square error (MSE) in the 3 ROIs. The influx rate maps calculated by WLSCF and our method are shown in Fig.4(b) and (c). Our method generates parametric maps with better contrast and MSEs are overall stable in 3 ROIs.

4 Conclusion

A robust kinetic parameter estimation framework is presented, which is robust to both poor statistical properties in dynamic PET and uncertainties in arterial input function, and is able to analyze every single kinetic parameter.

Acknowledgement. This work is supported in part by the National Basic Research Program of China (No: 2010CB732504) and by the Department of Science and Technology of Zhejiang Province(No: 2010C33026).

References

1. Bailey, D., Townsend, D., Valk, P., Maisey, M.: Positron Emission Tomography: Basic Sciences. Springer, Heidelberg (2005)
2. Visser, E., Philippens, M., Kienhorst, L., Kaanders, J.: Comparison of tumor volumes derived from glucose metabolic rate maps and SUV maps in dynamic 18F-FDG PET. Journal of Nuclear Medicine 49(6), 892 (2008)
3. Tsoumpas, C., Turkheimer, F., Thielemans, K.: A survey of approaches for direct parametric image reconstruction in emission tomography. Medical Physics 35, 3963 (2008)
4. Kamasak, M.E., Bouman, C.A., Morris, E.D., Sauer, K.: Direct reconstruction of kinetic parameter images from dynamic pet data. IEEE Transactions on Medical Imaging 25, 636–650 (2005)
5. Wang, G., Qi, J.: Generalized algorithms for direct reconstruction of parametric images from dynamic PET data. IEEE Transactions on Medical Imaging 28(11), 1717–1726 (2009)
6. Gunn, R.: Tracer Kinetic Modeling via Basis Pursuit. In: Senda, M., Kimura, Y., Herscovitch, P. (eds.) Brain Imaging using PET. Academic Press, New York (2002)
7. Cobelli, C., Foster, D., Toffolo, G.: Tracer Kinetics in Biomedical Research: From Data to Model. Kluwer Academic/Plenum Publishers, New York (2000)
8. Gao, F., Liu, H., Shi, P.: Efficient Robust Reconstruction of Dynamic PET Activity Maps with Radioisotope Decay Constraints. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6363, pp. 571–578. Springer, Heidelberg (2010)
9. Simon, D.: Optimal state estimation: Kalman, H $_\infty$ and nonlinear approaches. John Wiley and Sons, Chichester (2006)

Synthetic Echocardiographic Image Sequences for Cardiac Inverse Electro-Kinematic Learning

Adityo Prakosa[1], Maxime Sermesant[1], Hervé Delingette[1], Eric Saloux[2], Pascal Allain[3], Pascal Cathier[3], Patrick Etyngier[3], Nicolas Villain[3], and Nicholas Ayache[1]

[1] Asclepios Research Project, INRIA Sophia-Antipolis, France
[2] Service de Cardiologie CHU Caen, France
[3] Medisys, Philips Healthcare Suresnes, France

Abstract. In this paper, we propose to create a rich database of synthetic time series of 3D echocardiography (US) images using simulations of a cardiac electromechanical model, in order to study the relationship between electrical disorders and kinematic patterns visible in medical images. From a real 4D sequence, a software pipeline is applied to create several synthetic sequences by combining various steps including motion tracking and segmentation. We use here this synthetic database to train a machine learning algorithm which estimates the depolarization times of each cardiac segment from invariant kinematic descriptors such as local displacements or strains. First experiments on the inverse electro-kinematic learning are demonstrated on the synthetic 3D US database and are evaluated on clinical 3D US sequences from two patients with Left Bundle Branch Block.

1 Introduction

Despite advances in both medical image analysis and intracardiac electrophysiological mapping technology, the understanding of the relationship between the cardiac electrophysiology and the cardiac motion visible in images is only partial. However such understanding would be very valuable as it would open possibilities in non-invasive electrophysiological mapping. Since 3D echocardiography (US) is readily available, an important topic of interest for cardiologists would be the estimation of the cardiac electrophysiology function from the analysis of 3D US images. This is specifically important, for example, in the evaluation of the Cardiac Resynchronization Therapy (CRT) where the placement and tuning of pacemaker leads play a crucial role in the outcome of the therapy. In this context, cardiologists need to interpret time series of US images in order to detect and characterize kinematic patterns (motion asynchrony, delayed contraction) and then infer possible electrical conduction disorders.

While there is an important literature on the estimation of the cardiac kinematics from 3D US sequences (see for instance [3] and references therein), there exists no such tools to estimate the electrical wave propagation from such image sequences. However, the relationship between cardiac motion and electrical activation has been investigated in several studies [5–7].

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 500–507, 2011.

In this paper, we propose to study the inverse electro-kinematic relationship through the creation of a large database of synthetic 3D US images. Because it is difficult to obtain a large number of cases where both electrophysiological mapping and 3D US images are available, we use an electromechanical (E/M) model of the heart to produce synthetic but realistic image sequences for which the electrical stimulation is known. Previous work [6, 7] has mainly focused in detecting E/M wave directly from the displacement and strain patterns estimated from image sequences during the contraction and relaxation of the myocardium. Since the relationship between those mechanical waves and electrical waves is certainly complex, our approach is to learn it through an E/M model of the heart. Compared to [5], instead of estimating displacements and strains from the E/M model, we propose a more realistic estimation by first simulating 3D US images and then using an image-based motion tracking algorithm. Furthermore, rather than learning the activation forces over time, we have chosen to learn the depolarization times of all American Heart Association (AHA) segments. Finally, our learning approach is optimized in order to detect which kinematic descriptor is most correlated with the electrophysiology waves.

Different studies have been conducted for the creation of simulated 3D US sequences, e.g. [2, 3]. Instead of simulating the ultrasonic image formation process, in this paper, we propose a new approach to create synthetic 3D US sequences by deforming a real 3D US sequence and combining simulated myocardium displacements with the visible motion of the surrounding environment (blood pool speckle, mitral valve). This approach has the advantage of providing a realistic 3D US sequence at little computational cost and including all neighboring structures. A vast database of electrical propagations along with corresponding synthetic sequences based on the E/M simulation was created. On this database, invariant kinematic descriptors were extracted from each synthetic sequence and then fed to a machine learning algorithm which estimates the electrical pattern from kinematic descriptors during the cardiac cycle. The created synthetic 3D US sequences are of realistic quality and first experiments on the inverse electro-kinematic learning using this database are discussed.

2 Creating Synthetic 3D US Sequences

2.1 3D US Sequence Non-rigid Registration

We use as input to our method a real 3D US sequence acquired by the iE33 Philips probe on a patient suffering from heart failure. The first step in the pipeline was to segment semi-interactively or automatically the left ventricle (LV). The binary mask was then used to apply the iLogDemons non-rigid registration algorithm [4] which had been applied in the cardiac cine MR sequence analysis. This motion tracking algorithm enforces the incompressibility of the myocardium during the cardiac motion which provides an additional prior information to regularize the visible motion in the image sequence. With this non-rigid registration algorithm, the displacement field (DF) u between the end diastole (ED) image and each image of the real 3D US sequence was estimated

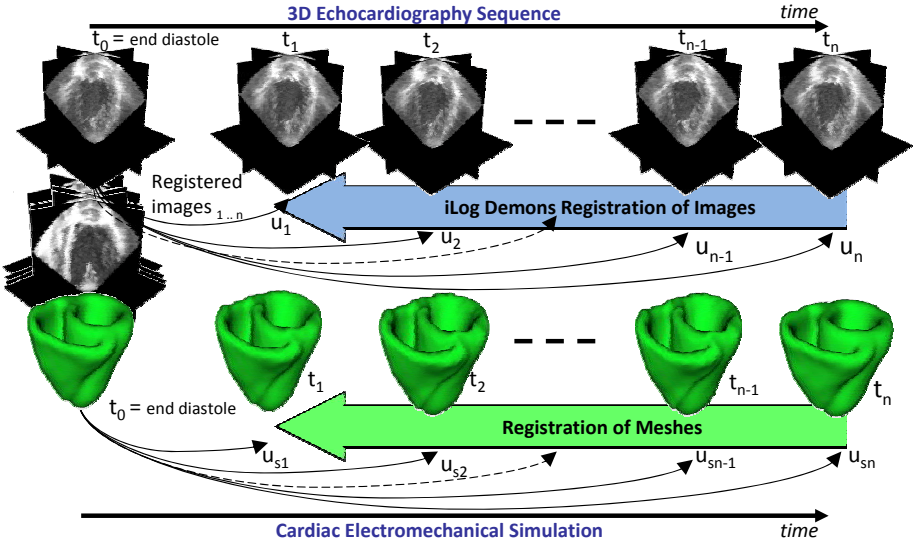


Fig. 1. Registration of Images and Meshes. iLog Demons registration method is applied to all images in the sequence to register them to the ED reference image. All meshes in a simulation cycle are also registered to the ED mesh.

(see Fig. 1). Thanks to the diffeomorphic nature of u, we computed its inverse and thus resampled each image of the sequence in the ED geometry.

2.2 Deformation of Registered 3D US Images Using E/M Simulation

From the segmented images of the myocardium at ED, we created a computational tetrahedral mesh which was suitable for the simulation of a cardiac E/M model [8] whose myocardium motion is used for the generation of the synthetic sequences. This required additional work since only part of the LV and right ventricle (RV) were visible in the image. Registration of a template mask of the 2 ventricle was used to infer the missing parts.

With this model, we simulated the cardiac motion after specifying an electrophysiological pattern (see Section 2.3). We sampled the cardiac simulated motion to follow the temporal resolution of the real 3D US sequence and then computed the DF between the reference configuration (ED) and the deformed position at each time of the sequence using the linear interpolation of the displacement of each vertex of the tetrahedral mesh rasterized in a 3D image having the same size and spatial resolution as the real 3D US image (see Fig. 2). This dense synthetic DF of the myocardium was then merged with the DF estimated from the non-rigid registration. The synthetic DF completely overwrites the registration DF within the myocardium. Additionally, the synthetic DF within the eroded myocardium is diffused by solving the Laplace equation and fused with

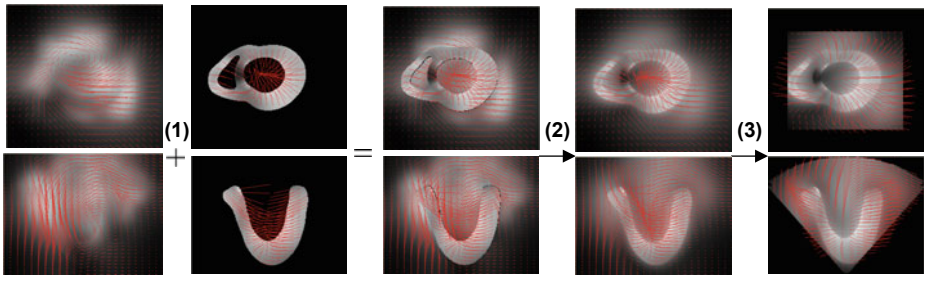


Fig. 2. Fusion of the Displacement Fields. (1) The DF estimated from the iLog Demons registration (left) is combined with the myocardium DF from the E/M simulation (second left). The two fields are fused, (2) smoothed, (3) inversed and cropped along the acquisition cone.

the registration DF to smooth the transition outside the myocardium. Then, the new DF was inversed and applied to each real image previously resampled in the ED configuration. Finally, a 3D cone mask was applied to remove all the displacements outside the cone, as observed in real acquisitions. With this approach, most of the image will stay unchanged in the synthetic image compared to the original sequence. We preserve the dynamics of the image, in particular the speckle visible in 3D US for most voxels. Only in the myocardium is the image texture slightly warped, the amount of warping depending on the difference between the simulated cardiac motion and the motion in the original images.

2.3 Generation of Healthy and Pathological Cardiac Motion

Different simulation scenarios were performed including normal and pathological cases such as left bundle branch block (LBBB) and right bundle branch block (RBBB) by blocking the LV and RV initial electrical activation respectively, LBBB with LV pacing, RBBB with RV pacing and also LBBB and RBBB with biventricular (BV) pacing. The different pacing positions were based on the LV AHA segments (see Fig. 3). Table 1 summarizes the electrical and mechanical parameters used for the 120 simulations done from each real 3D US sequences.

3 Learning Electro-Kinematic Inverse Relationship

3.1 Kinematic Descriptors

With the method described previously, a large database of synthetic 3D US images was created. We then tracked the cardiac motion from those synthetic images by using the iLogDemons registration algorithm [4]. More precisely, we registered all the images of the synthetic sequence to its reference ED image. As an input to a machine learning algorithm, we needed to first extract kinematic descriptors which describe in a compact and exhaustive way the cardiac motion.

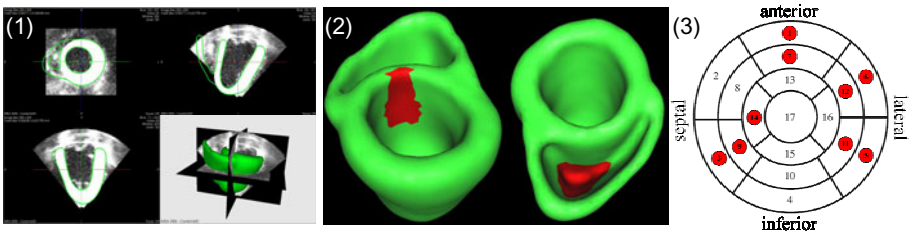


Fig. 3. Cardiac Geometry and Electrical Stimulation. (1) LV segmentation (2) Initial electrical activation area for the normal stimulation (3) Positions of the stimulation leads in the LV AHA zones

Table 1. Simulation Database. Parameters of the 120 simulations. Global conductivity (cm/s) is the conduction velocity of the electrophysiology model and global contractility (adimensioned) is the peak contractility of the E/M coupling.

Simulation Number	Initial Electrical Activation Position	Global Conductivity	Global Contractility
1-4	LVRV (Normal)	50/30	0.09/0.05
5-8	LV (RBBB)	50/30	0.09/0.05
9-12	RV (LBBB)	50/30	0.09/0.05
13-36	RV + AHA 1/5/6/7/11/12 (LV Pacing)	50/30	0.09/0.05
37-48	LV + AHA 3/9/14 (RV Pacing)	50/30	0.09/0.05
49-120	AHA 1/5/6/7/11/12 + AHA 3/9/14 (BV Pacing)	50/30	0.09/0.05

To this end, we characterized the motion of each AHA segment by fitting in the least-square sense an affine transformation $f(p) = Ap + B$ to the iLogDemons estimated DF. The strain tensor was computed from the affine matrix $E = \frac{1}{2}(A^T A - I)$. We propose to extract kinematic descriptors that are invariant to any change of reference frame (or rigid transformation). For the strain matrix E, the three Euclidean invariants are written as $x_1 = \text{trace}(E)$, $x_2 = \text{trace}(E^2)$, and $x_3 = \det(E)$. For the displacement vector, we only extracted its norm as invariant: $x_4 = \|u\| = \|Ab + B - b\|$, where $\|u\|$ is the displacement norm of the zone centroid with b the initial position of the centroid. Finally, we also used the strain in the direction of displacement as the last invariant $x_5 = \frac{1}{2\|u\|^2}(u^T E u)$. These 5 descriptors for the 17 AHA zones during the 19 time instances of a cardiac cycle were used to create a vectorial kinematic descriptor for each simulation: $X = x_i \in \mathbb{R}^d$ where d=5 (Descriptors) $\times$ 19 (Times) $\times$ 17 (Zones) = 1615.

3.2 Inverse Electro-Kinematic Learning

In the inverse electro-kinematic learning process, the non-linear relationship between the kinematic descriptors and the electrical propagation was estimated based on a training set extracted from the synthetic database. To represent the

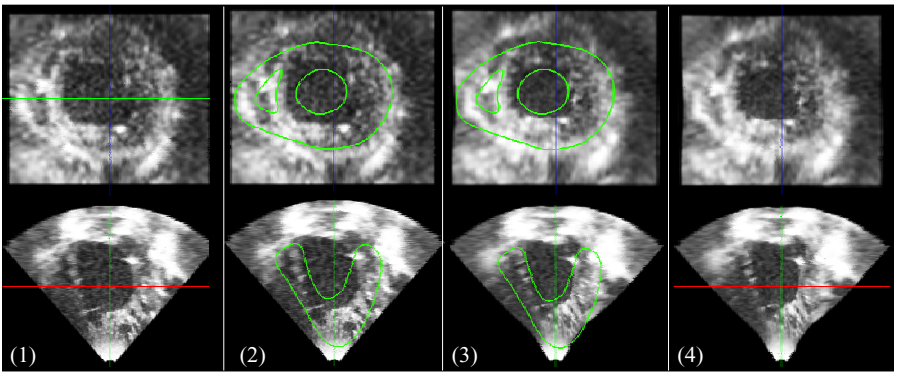


Fig. 4. Synthetic 3D US. (1) original real image with (2) contour of the mesh at the corresponding time from the model simulation overlayed, (3) synthetic image generated with the model simulation with model contour overlay, (4) synthetic image.

cardiac electrophysiology, we considered the activation time when the electrical potential starts to depolarize at a point of the myocardium. The activation time was averaged for all points in each AHA segment. Therefore, the vector characterizing electrophysiology for each simulation is $\mathrm{Y} = \mathrm{y}_i \in \mathbb{R}^{\mathrm{r=17\ (AHA\ Zones)}} = log(\mathrm{Activation\ Times})$.

We modelled the non-linear relationship using Least-Square Support Vector Machine (LS-SVM) $Y = f(X) = Ak(x_i, X) + b$ with the Radial Basis Function (RBF) $K(x_i, x_j) = e^{-z}$ as the Kernel function where $z = \sum_{k=1}^{5} \left(\frac{|x_i^k - x_j^k|}{\sigma_k \alpha_k} \right)^2$. In this kernel function, σ_k is the standard deviation of each descriptor and α_k is a dimensionless coefficient which weights the importance of the descriptor in the learning process. Finally, following the LS-SVM theory, $k(x_i, X)$ is a kernel vector while matrix A is computed as $\mathrm{A} = Y^T (\lambda \mathrm{I} + K)^{-1}$. In order to have a good generalization of the model, the α_k parameters and the regularization parameter λ were optimized with a downhill simplex method using leave-one-out cross-validation based on Allen's predicted residual sum-of-squares (PRESS) statistic [1].

4 Results

The proposed synthetic 3D US generation method produces realistic synthetic 3D US sequence (cf. Fig. 4) with a seamless fusion of simulated myocardium motion with neighboring moving structures. The created synthetic 3D US database contains 120 different cardiac cases consisting of a sequence of 19 3D US images describing a complete cardiac cycle. In total, $120 \times 19 = 2280$ synthetic 3D US images were generated.

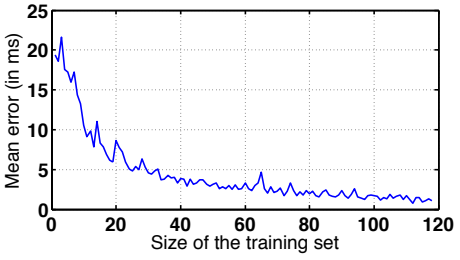


Fig. 5. RMS Residual vs Size of Training Data. Less than 10 ms RMS residual is obtained by using more than 15 training cases.

4.1 Machine Learning Validation on Synthetic Data

We evaluated the learning process on synthetic data and estimated the minimum size of the training set to have a small regression error for the remaining entries of the database. Fig. 5 shows a good generalization with a root mean square (RMS) error of less than 10 ms of residual by using at least 15 training datasets.

4.2 Machine Learning Evaluation on Real Data

After optimizing the PRESS criterion on the whole synthetic database, we obtained the following LS-SVM parameters : $\lambda = 7.89 \times 10^{-31}$, $\alpha_1 = 463.65$, $\alpha_2 = 2.29 \times 10^{13}$, $\alpha_3 = 8.02 \times 10^{12}$, $\alpha_4 = 14.37$ and $\alpha_5 = 174.51$. This clearly shows that the kinematic descriptors x_1, x_4 and x_5 are the only meaningful ones to learn the electro-kinematic relationship. We did a first evaluation of this learning process on clinical 3D US sequences for two patients with LBBB. After performing non-rigid registration and extracting the vector X of kinematic descriptors, the electrophysiology vector Y was estimated from the LS-SVM. Very similar estimated depolarization times were obtained for these two patients (cf. Fig. 6). Moreover, the activation patterns correspond to what was expected: depolarization starts from the septum towards the lateral wall, and the difference between the first activated zone and the last activated zone, which indicates the QRS duration, is around 150 ms which is also a characteristic of the LBBB.

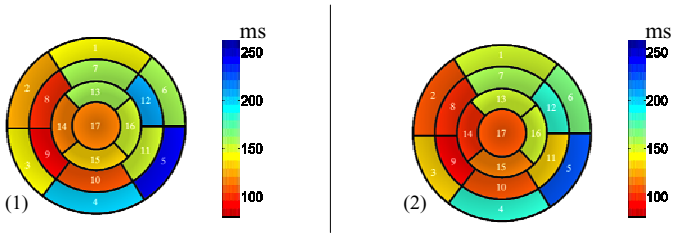


Fig. 6. Depolarization Time Estimation from Clinical 3D US Sequences. First evaluation of the learning process on patient (1) and patient (2). Both patients have LBBB.

5 Conclusion

We developed a pipeline to create realistic synthetic 3D US sequences using the deformation from an E/M model simulation. Those sequences represent in themselves a valuable result for instance to benchmark motion tracking algorithms. As these synthetic 3D US sequences have electro-kinematic "ground truth" information, we thus performed an inverse electro-kinematic learning on this database. Invariant kinematic descriptors were extracted from the DF obtained from the synthetic 3D US images registration. The non-linear inverse relationship between the electrical activation times and the kinematic descriptors was modelled using LS-SVM. Evaluation of the learning process for the synthetic 3D US sequences database shows good generalization and the first evaluation on clinical 3D US sequences shows encouraging results.

Acknowledgement. This work was partially supported by the Care4me ITEA2 project and the European Community Seventh Framework Programme (FP7/ 2007- 2013) under grant agreement n 224495 (euHeart project).

References

1. Cawley, G.C.: Leave-one-out cross-validation based model selection criteria for weighted LS-SVMs. In: IJCNN 2006, pp. 1661–1668. IEEE, Vancouver (2006)
2. Duan, Q., Moireau, P., Angelini, E.D., Chapelle, D., Laine, A.F.: Simulation of 3D ultrasound with a realistic electro-mechanical model of the heart. In: Sachse, F.B., Seemann, G. (eds.) FIHM 2007. LNCS, vol. 4466, pp. 463–473. Springer, Heidelberg (2007)
3. Elen, A., Choi, H.F., Loeckx, D., Gao, H., Claus, P., Suetens, P., Maes, F., D'hooge, J.: Three-dimensional cardiac strain estimation using spatio-temporal elastic registration of ultrasound images: A feasibility study. IEEE Trans. Med. Imaging 27, 1580–1591 (2008)
4. Mansi, T., Pennec, X., Sermesant, M., Delingette, H., Ayache, N.: iLogDemons: A demons-based registration algorithm for tracking incompressible elastic biological tissues. Int. J. of Comput. Vision 92, 92–111 (2011)
5. Prakosa, A., Sermesant, M., Delingette, H., Saloux, E., Allain, P., Cathier, P., Etyngier, P., Villain, N., Ayache, N.: Non-invasive activation times estimation using 3D echocardiography. In: Camara, O., Pop, M., Rhode, K., Sermesant, M., Smith, N., Young, A. (eds.) STACOM 2010. LNCS, vol. 6364, pp. 212–221. Springer, Heidelberg (2010)
6. Provost, J., Lee, W., Fujikura, K., Konofagou, E.: Electromechanical wave imaging of normal and ischemic hearts in vivo. IEEE Trans. Med. Imaging 29, 625–635 (2010)
7. Sanchez-Ortiz, G., Sermesant, M., Chandrashekara, R., Rhode, K., Razavi, R., Hill, D., Rueckert, D.: Detecting the onset of myocardial contraction for establishing inverse electro-mechanical coupling in XMR guided RF ablation. In: ISBI 2004, pp. 1055–1058. IEEE, Arlington (2004)
8. Sermesant, M., Delingette, H., Ayache, N.: An electromechanical model of the heart for image analysis and simulation. IEEE Trans. Med. Imaging 25, 612–625 (2006)

Automatic Prone to Supine Haustral Fold Matching in CT Colonography Using a Markov Random Field Model

Thomas Hampshire[1], Holger Roth[1], Mingxing Hu[1], Darren Boone[2], Greg Slabaugh[3], Shonit Punwani[2], Steve Halligan[2], and David Hawkes[1]

[1] Centre for Medical Image Computing, University College London, London, UK
[2] Department of Specialist Radiology, University College Hospital, London, UK
[3] Medicsight PLC, London, UK

Abstract. CT colonography is routinely performed with the patient prone and supine to differentiate fixed colonic pathology from mobile faecal residue. We propose a novel method to automatically establish correspondence. Haustral folds are detected using a graph cut method applied to a surface curvature-based metric, where image patches are generated using endoluminal CT colonography surface rendering. The intensity difference between image pairs, along with additional neighbourhood information to enforce geometric constraints, are used with a Markov Random Field (MRF) model to estimate the fold labelling assignment. The method achieved fold matching accuracy of 83.1% and 88.5% with and without local colonic collapse. Moreover, it improves an existing surface-based registration algorithm, decreasing mean registration error from 9.7mm to 7.7mm in cases exhibiting collapse.

1 Introduction

Computed tomographic colonography (CTC) is widely considered the preferred radiological technique for detecting colorectal cancer or potentially precancerous polyps. When characterising potential polyps, the radiologist must manually match corresponding areas in the prone and supine data. However, this is a difficult, time-consuming task due to considerable deformation that occurs during repositioning [1]. Hence, a method for automatic registration of prone and supine datasets has the potential to improve radiologists' efficiency and confidence. Furthermore, accurate surface registration could improve specificity of computer-aided detection systems (CAD).

A number of methods have been proposed to find correspondence between the prone and supine positions. For example, centreline-based methods extract and align colonic centrelines by stretching and shrinking based on path geometries [2]. Anatomical landmarks can be used to help align the two datasets by first identifying a stable set of anatomical features, such as the caecum, rectum and flexures [3]. Voxel-based methods provide a further means of registration [4]. However, these methods rely to varying extents upon continuous prone-supine

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 508–515, 2011.

colonic segmentations, free from occlusion by fluid or collapse; a scenario which occurs infrequently in daily practice, despite optimal bowel preparation [5].

Fukano et al. proposed a registration method based on haustral fold matching [6]. A second-order derivative difference filter was used to extract folds; their volume and relative positions along the centreline were used for matching. The method relied on prior automatic identification of a set of landmark locations for registration. They reported correct registration of only 65.1% of large folds and 13.3% of small folds.

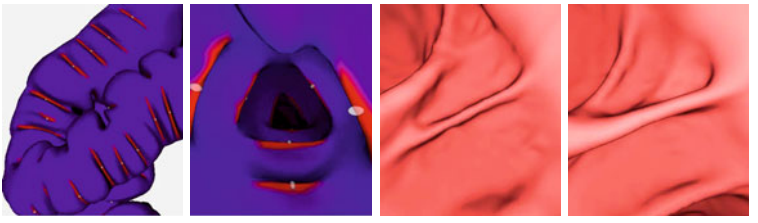
Recently, methods which involve conformal mapping of the colonic surface have been proposed in order to reduce the complexity of the three-dimensional task. For example, Zeng et al. combined conformal mapping with feature matching between the prone and supine surfaces [7]. The prone and supine colonic segmentations were mapped onto five rectangle pairs. Correspondences were established using a feature matching method based upon mean curvature. The method relied on accurately determining five matching segments in the prone and supine datasets, which is difficult to achieve and may not be possible in the case of local colonic collapse. The method proposed by Roth et al. [8], aims to overcome these limitations by mapping the entire endoluminal surface to a cylinder. Dense surface correspondence was then achieved by non-rigid cylindrical B-spline registration, driven by local shape measurements. However, this method can be susceptible to mis-registration of continuous sections due to the similarities of neighbouring features.

We present a novel method for generating a set of robust landmark correspondences between the prone and supine CT data. While previous methods tried to match corresponding folds based on spatial location and size alone, e. g. [6,7], we also compare visual renderings of the colonic surface at the fold positions as well as local geometric information, without reliance on a conformal mapping. The matching problem is modelled with an MRF and the maximum a posteriori labelling solution is estimated to provide a correspondence. This method explicitly addresses the problem of colonic collapse and can provide an initialisation in order to improve methods which aim to provide a full surface correspondence.

2 Methods

2.1 Haustral Fold Segmentation

Haustral folds are elongated, ridgelike structures on the endoluminal surface which can be identified by extracting curvature measurements from a surface reconstruction of the colonic wall. The maximum and minimum values of the normal curvature at a point are called the principal curvatures, k_1 and k_2 repectively. A metric based on the principal curvatures is used to classify each vertex as fold, or non-fold: $M = k_1 - \gamma||k_2||$. This recognises that at a fold, one expects $k_1 >> 0$ and $k_2 \approx 0$. The γ parameter penalises the metric against curvature in any direction other than in the maximum, helping to separate the folds at the tenaie coli. The surface mesh is treated as a graph, with graph nodes defined



(a) (b) (c) (d)

Fig. 1. External (a) and internal (b) views of segmented haustral folds with marked centres. Virtual colonoscopy views of corresponding folds in the prone (c) and supine (d)

by the mesh vertices and graph edges defined by the mesh edges. Using a virtual sink and source with the given weighting, a graph cut segmentation [9] is performed which minimises an energy function using M as a unary term and a Potts energy function smoothing term: $\delta(f_p, f_q) = (f_p == f_q) \cdot \zeta$. This results in a label assignment of fold or non-fold over the entire surface mesh (Fig. 1). The centre of each fold is taken as the vertex with the shortest maximum distance to any vertex lying on the border of the segmented region. An observer study was carried out using 4 data sets to evaluate the segmentation performance. Given these cases, a sensitivity of 86.5% and positive predictive value of 97.0% is achieved.

2.2 Markov Random Field Modelling

The matching of prone and supine haustral folds is formulated as a labelling problem. First, m haustral folds are detected in the supine data and these are uniquely labelled; the objective is to then assign labels to the detected prone folds, achieved by solving an MRF. The identified haustral folds in the prone data set are modelled as sites $\mathbf{S} = S_1, ..., S_n$, each of which has an associated random variable $\mathbf{F} = F_1, \ldots, F_n$ taking on a discrete label $\mathbf{f} = f_1, \ldots, f_n$ taken from the set of haustral folds identified in the supine data set. A neighbourhood system $\mathbf{N} = N_i | \forall i \in \mathbf{S}$ defines the extent of local connections between sites; and a pair-wise clique defined on $\mathbf{N}$ and $\mathbf{S}$, $\mathbf{C}_2 = \{i, i^{'}\} | N_i, i \in \mathbf{S}$ allows the incorporation of a-priori knowledge of geometric dependencies between labels.

The maximum a posteriori (MAP) estimate of the optimum labelling is computed, which is equivalent to minimising the energy function:

$$\mathbf{f}^{(MAP)} = \arg\min_{\mathbf{f}} \left[\sum_{i \in \mathbf{S}} V_u(f_i) + \sum_{i \in \mathbf{S}} \sum_{j \in N_i} V_p(f_i, f_j) \right] \tag{1}$$

where $V_u(f_i)$ is the unary term, a cost function for assigning label value f_i to site S_i. The pair-wise term $V_p(f_i, f_j)$ is the cost for assigning neighbouring sites S_i and S_j to their current values.

Unary Cost Function. The aim is to calculate an $n \times m$ unary cost matrix, where $n = \|\mathbf{S}\|$ is the number of sites or prone folds, and $m = \|\mathbf{L}\|$ is the number of labels or supine folds. To calculate the cost $V_u(f_i = L_j)$ of assigning label L_j to site S_i image patches are rendered at the fold positions, visualising the internal colon wall (Fig. 1). The resulting images are then compared using a sum-of-squared-difference similarity metric. An optimisation over the external parameters of the virtual camera used to visualise the supine dataset accounts for any inaccuracies in the fold point identification. Restricting the number of degrees of freedom of camera search ensures that the camera focus remains on the correct fold. The degrees of freedom are as follows: Elevation (θ) - the fold centre and camera right vector give a position and axis about which the the camera is rotated; roll (ϕ) - rotation around the camera view direction; dollying (τ) - translation along the camera right vector.

Given the three parameters θ, ϕ, τ; the optimisation finds the local minimum in a mean sum of squared difference between the rendered images I_1 and I_2 using Powell's gradient descent method [10]. Adding a scaling parameter W allows the weighting of unary to pair-wise costs. Additionally, a constraint is added so that the matching folds must lie in a similar region. The fold centreline positions ν_s^c, ν_p^c are used to limit corresponding fold matches to a window of one fifth of the colon length. Finally, a constant unary cost α is associated with the assignment of the null label to any given node, allowing for missing labels. We define $R(I_1, I_2) = W\sqrt{MSSD(I_1, I_2|\theta, \phi, \tau)}$. The unary costs are then defined:

$$V_u(f_i = L_j) = \begin{cases} \alpha & \text{if } L_j = L_0 \\ R(I_1, I_2) - \min(R(I_1, \cdot)) & \text{if } \nu_{s\to p}^c \leq \|\nu_{full}^c\|/10 \\ \infty & \text{otherwise} \end{cases} \quad (2)$$

Pair-wise Cost Function. To improve labelling performance, geometric information about neighbouring fold positions can be used. In this work a Rotation Minimising Frame (RMF) [11] is employed to describe the relative position of each fold to its neighbours: $\boldsymbol{\nu} = [\nu^c, \nu^\theta]^T$; where ν^c is the difference in fold position along the centreline and ν^θ, the difference in angle of rotation around the centreline. This 2D parameterisation simplifies the description of the translation between corresponding pairs of folds between the prone and supine as the centreline ν^c and rotational ν^θ displacement should be similar ($\boldsymbol{\nu}_p \approx \boldsymbol{\nu}_s$). Alternatively we can state $\boldsymbol{\nu}_p = \boldsymbol{\nu}_s + \epsilon$, where ϵ represents some uncertainty, and can be modelled with a zero mean bivariate normal distribution $\epsilon \sim \mathcal{N}(0, \boldsymbol{\Sigma})$, with $\boldsymbol{\Sigma} = diag(\Sigma^c, \Sigma^\theta)$. Finally we recognise that the position of a neighbouring site becomes more uncertain as the displacement along the centreline increases. For each site pair $\{i, j\}$ we calculate $\boldsymbol{\Sigma}_{ij} = (I + (\nu_S^c \cdot \boldsymbol{\lambda}))\boldsymbol{\Sigma}_{base}$, where $\boldsymbol{\lambda} = diag(\lambda_1, \lambda_2)$ is a parameter to control the increase in positional uncertainty with centreline displacement, and $\boldsymbol{\Sigma}_{base}$ is a base covariance. With this information, a pair-wise cost for assigning neighbouring sites S_i and S_j label configurations f_i and f_j, is defined by the negative log-normal distribution: $\varphi(f_i, f_j) = (\boldsymbol{\nu}_S - \boldsymbol{\nu}_L)^T \boldsymbol{\Sigma}_{ij}^{-1}(\boldsymbol{\nu}_S - \boldsymbol{\nu}_L)/2$. A local neighbourhood system is defined in order to enforce local geometric constraints on neighbouring fold positions.

The local neighbourhood of a site is set to be $N_i^{local} = \{i \neq j \mid \|\nu_i^c - \nu_j^c\| < t\}$, where t is a threshold distance. Pairs of sites that are separated by a local colonic collapse are removed from the neighbourhood set. A uniqueness constraint is also enforced so any two sites may not be assigned the same label. This is included in the pair-wise cost function by connecting each site with every other site in a global neighbourhood system $N_i^{global} = \{i \neq j\}$ and defining the pair-wise cost of assigning the same label to two nodes to be infinity, except in the case of a null label assignment $V_b(f_i, f_j) = \infty$ if $f_i = f_j \neq L_0$. The full pair-wise cost function is:

$$V_b(f_i, f_j) = \begin{cases} \infty & \text{if } f_i = f_j \\ \beta & \text{if } S_j \in N_i^{local} \text{ and } (f_i = L_0 \text{ or } f_j = L_0) \\ \varphi(f_i, f_j) & \text{if } S_j \in N_i^{local} \text{ and } f_i \neq L_0 \text{ and } f_j \neq L_0 \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

MRF Inference. The uniqueness constraint on the pair-wise costs means the problem of solving the MRF is non-submodular and restricts the possible algorithm choice for MAP inference. The Belief Propagation (BP) algorithm is suitable for this purpose. To estimate the maximum probability state configuration, the min-sum variant is used [12]. Parameters are found using a gradient ascent optimisation on training data.

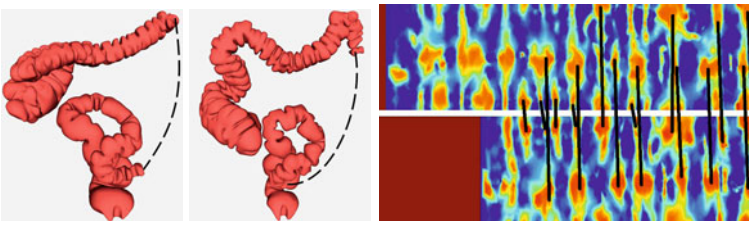
3 Experimental Results

3.1 Clinical Validation

Ethical approval and informed consent was obtained to use anonymised CT colonography data. Colonic cleansing and insufflations had been performed in accordance with current recommendations [13]. A radiologist (with experience of over 500 endoscopically validated studies) used virtual colonoscopic reconstructions to identify corresponding folds in the prone and supine datasets to establish a reference standard. Any folds where a confident manual correspondence could not be established were disregarded. This resulted in a total of 1175 corresponding fold pairs over 13 datasets, 5 of which exhibited at least one local colonic collapse in one or both views (case 10 is shown in Figure 2). For a small subset of 3 randomly selected cases the reference standard was reestablished after a period of three months. Folds which have a correspondence in both sets of reference standard are used to evaluate the level of intra-observer variability, showing a 85.3% agreement.

3.2 Haustral Fold Matching

To assess the performance of the algorithm, for each case the maximum a posteriori labelling solution is compared against the reference standard described above. Table 1 shows the results for the cases with and without colonic collapse. Although the percentage of correctly labelled folds is high, at 83.1% and 88.5%



(a) (b) (c)

Fig. 2. External view of prone (a) and supine (b) datasets in case 10. The dotted line indicates an area of colonic collapse. (c) shows the conformally mapped colonic prone (top) and supine (bottom) surfaces at the sigmoid with shape index values. The rectangles represent unfolded cylinders. Red areas indicate a collapsed region. Black lines show detected fold correspondences of which all agree with the reference standard.

Table 1. Fold labelling performance. The number of Reference Standard (RS) points are shown. Label shows the percentage of RS points identified with a label.

	Without colonic collapse									With colonic collapse					
Case	1	2	3	4	5	6	7	8	Total	9	10	11	12	13	Total
RS Points	74	104	112	88	86	112	107	91	774	65	107	66	83	80	401
Labelled	66	97	106	84	82	92	99	88	714	62	101	63	77	51	354
Correct	49	90	98	70	74	76	91	84	632	50	78	53	74	39	294
Incorrect	17	7	8	14	8	16	8	4	82	12	23	10	3	12	60
Label(%)	89.2	93.3	94.6	95.5	95.3	82.1	92.5	96.7	92.2	95.4	94.4	95.5	92.8	63.8	88.3
Correct(%)	74.2	92.8	92.5	83.3	90.2	82.6	91.9	95.5	88.5	80.6	77.2	84.1	96.1	76.5	83.1

for cases with and without colonic collapse respectively, it is apparent that some cases show a much higher rate of accuracy than others. This is primarily due to different levels of distension causing inconsistent fold identification between data sets. Whilst the method is robust to missing data, a greater proportion of correctly labelled folds is demonstrated in cases with better fold segmentation. A second factor which contributes largely to the variance in fold pair matching in the pair-wise term, is the inaccurate estimation of the angle of rotation between folds using the centreline-based RMF estimation method, especially apparent in areas of high torsion, such as the flexures.

3.3 Initialisation of Surface Based Registration Method

There are scenarios where obtaining a one-to-one surface correspondence is required. In this case, the results of this fold matching method can be used to provide automated initialization for a surface-based registration technique [8]. Here, the fold positions are mapped onto a conformally mapped image and used to provide a linear scaling between haustral folds in the direction of the centreline, prior to the B-spline registration where full surface correspondence is established. To determine the registration error, each reference standard point is

Table 2. Surface registration initialisation with non-collapsed cases. The number of Reference Standard (RS) points are shown. Error 1 and 2 show the error of the surface-based registration without and with using points as an initialisation.

	Without colonic collapse									With colonic collapse					
Case	1	2	3	4	5	6	7	8	Total	9	10	11	12	13	Total
RS Points	74	104	112	88	86	112	107	91	774	65	107	66	83	80	401
Error 1 (mm)	11.5	8.6	5.3	5.7	5.5	5.2	5.8	6.7	6.6	12.2	6.5	7.8	13.5	9.6	9.7
Error 2 (mm)	11.5	7.2	5.5	5.7	5.8	5.5	6.1	6.9	6.6	7.9	5.8	7.8	8.7	9.1	7.7
Difference (mm)	0.0	-1.4	0.2	0.0	0.3	0.3	0.3	0.2	0.0	-4.3	-0.7	0	-4.8	-0.5	-2

transformed from one dataset to the other using the registration result, and the 3D Euclidean distance between this and the corresponding reference standard point is measured. The results for cases with and without colonic collapse are shown in Table 2 using the same reference standard as in the previous experiment. It can be seen that the initialisation improves registration in cases exhibiting local colonic collapse, decreasing the mean error from $9.7mm \pm 8.7mm$ to $7.7mm \pm 7.1mm$; however in cases without a local colonic collapse, the mean error was unchanged at $6.6mm$. This shows that the fold matching technique is more robust than the surface-based registration in the case of poor insufflation (e.g. collapse); however, in well distended cases, the space to improve upon the registration of the surface-based method is limited. Using a Related Samples Wilcoxon Signed Rank Test with a significance level set at $p < 0.01$, the differences in mean errors are not significantly significant in the cases without colonic collapse ($p = 0.317$). However, in the cases with colonic collapse a statistical significance is observed ($p = 0.009$).

4 Conclusion and Future Work

Although several registration methods have been proposed, they often require manual initialisation using anatomical landmarks or are likely to fail in the presence of local colonic under-distension. Colonic collapse or segments with inadequate fluid tagging are very common in routine practice and methods must have the inherent capability to deal with this in order to be clinically useful. We propose a novel method for detecting and establishing correspondence between haustral folds in prone and supine CT colonography data sets which accurately deals with these issues. In addition, applying this method to initialise a surface-based registration technique can reduce registration error. It is clear that the unary costs are not as reliable in areas of high deformation due to the change in appearance of haustral folds. Investigation into a dynamic weighting of unary to pair-wise costs may allow for improved robustness in these situations. Future work would include the use of tenaie coli identification to provide a 'zero angle', improving the accuracy of the pair-wise costs. Although thorough clinical validation on a large dataset is ongoing, current results presented in this paper clearly demonstrate the promise of our MRF approach to match corresponding haustral folds in CTC data.

References

1. Punwani, S., Halligan, S., Tolan, D., Taylor, S.A., Hawkes, D.: Quantitative assessment of colonic movement between prone and supine patient positions during CT colonography. British Journal of Radiology 82(978), 475–481 (2009)
2. Wang, S., Yao, J., Liu, J., Petrick, N., Van Uitert, R.L., Periaswamy, S., Summers, R.M.: Registration of prone and supine ct colonography scans using correlation optimized warping and canonical correlation analysis. Medical Physics 36, 5595 (2009)
3. Näppi, J., Okamura, A., Frimmel, H., Dachman, A., Yoshida, H.: Region-based supine-prone correspondence for the reduction of false-positive cad polyp candidates in ct colonography. Academic Radiology 12(6), 695–707 (2005)
4. Suh, J.W., Wyatt, C.L.: Deformable registration of supine and prone colons for computed tomographic colonography. Journal of Computer Assisted Tomography 33(6), 902–911 (2009)
5. Taylor, S.A., Halligan, S., Goh, V., Morley, S., Bassett, P., Atkin, W., Bartram, C.I.: Optimizing Colonic Distention for Multi–Detector Row CT Colonography: Effect of Hyoscine Butylbromide and Rectal Balloon Catheter1. Radiology 229(1), 99 (2003)
6. Fukano, E., Oda, M., Kitasaka, T., Suenaga, Y., Takayama, T., Takabatake, H., Mori, M., Natori, H., Nawano, S., Mori, K.: Haustral fold registration in ct colonography and its application to registration of virtual stretched view of the colon. In: Proceedings of SPIE, vol. 7624, p. 762420 (2010)
7. Zeng, W., Marino, J., Gurijala, K.C., Gu, X., Kaufman, A.: Supine and prone colon registration using quasi-conformal mapping. IEEE Transactions on Visualization and Computer Graphics 16, 1348–1357 (2010)
8. Roth, H.R., McClelland, J.R., Boone, D.J., Modat, M., Cardoso, M.J., Hampshire, T.E., Hu, M., Punwani, S., Ourselin, S., Slabaugh, G.G., Halligan, S., Hawkes, D.J.: Registration of the endoluminal surfaces of the colon derived from prone and supine ct colonography. Medical Physics 38(6), 3077–3089 (2011)
9. Boykov, Y., Kolmogorov, V.: An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1124–1137 (2004)
10. Fletcher, R., Powell, M.J.D.: A rapidly convergent descent method for minimization. The Computer Journal 6(2), 163–168 (1963)
11. Klok, F.: Two moving coordinate frames for sweeping along a 3D trajectory. Computer Aided Geometric Design 3(3), 217–229 (1986)
12. Weiss, Y., Freeman, W.T.: On the optimality of solutions of the max-product belief-propagation algorithm in arbitrary graphs. IEEE Transactions on Information Theory 47(2), 736–744 (2002)
13. Taylor, S.A., Laghi, A., Lefere, P., Halligan, S., Stoker, J.: European society of gastrointestinal and abdominal radiology (esgar): consensus statement on ct colonography. European Radiology 17(2), 575–579 (2007)

Detecting Patient Motion in Projection Space for Cone-beam Computed Tomography

Wolfgang Wein and Alexander Ladikos

White Lion Technologies AG, Munich, Germany

Abstract. Cone-beam X-Ray systems strictly depend on the imaged object being stationary over the entire acquisition process. Even slight patient motion can affect the quality of the final 3D reconstruction. It would be desirable to be able to discover and model patient motion right from the actual projection images, in order to take it into account during reconstruction. However, while the source-detector arrangement is rotating around the patient, it is difficult to separate this motion from the additional patient motion. We present a novel similarity metric for successive X-Ray projections, which is able to distinguish between the expected rotational and additional sources of motion. We quantitatively evaluate it on simulated dental cone-beam X-Rays and qualitatively demonstrate its performance on real patient data.

1 Introduction

As opposed to rapidly spinning Computed Tomography (CT) devices, 3D cone-beam X-Ray systems take much longer (10-30 seconds) to acquire enough projection images for a 3D reconstruction. Both unexpected motion of the device itself (i.e. deviation from its calibrated circular path) as well as patient motion have negative effects on the reconstruction, if not modeled correctly. Depending on the organ being imaged, the latter can be categorized into respiratory movement, cardiac motion, as well as direct movement of the patient.

If additional information, such as an ECG sensor, is available which allows to separate motion phases, a binned reconstruction is possible, whereafter further motion refinement is possible [5]. In the absence of such information, both image and motion can be simultaneously reconstructed using an iterative framework [2]. While such approaches are computationally demanding, they potentially are able to recover small internal anatomic movement. However, convergence is not guaranteed if large-scale, non-periodic patient motion is present, as an initial reconstruction has to use either zero motion or a prior motion model.

Estimating motion in projection image space is difficult because the source-detector motion itself causes change. On top of that, the amount of change varies with the viewing angle and anatomy traversed by the X-Rays. In [6], an optical flow estimation between projection images and known reference images is used for a rough estimation. In [1] various standard similarity measures are evaluated with respect to their capability of distinguishing motion directly from successive X-Ray projections. Regardless of the specific measure used, such an approach can

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 516–523, 2011.

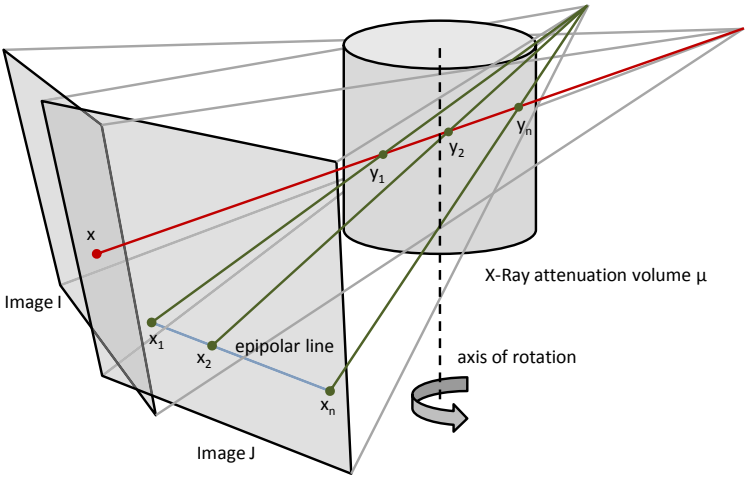


Fig. 1. Geometry of two successive X-Ray images during a cone-beam acquisition. For better illustration, the angle increment, and therefore the length of the epipolar line segment, has been increased.

only work if the amount of image change due to patient motion is significantly larger than the change induced by the device angle increment.

In the following, we present our approach for detecting motion in projection space. A new similarity measure is developed, which uses the projection geometry of two successive X-Ray images for estimating to which region in the second image a certain location in the first image can contribute. Inconsistent structural change can therefore be singled out and used to assess patient motion.

2 Methods

2.1 Geometric Modeling

Let us consider the geometric relationship of two successive cone-beam X-Ray projection images I and J as illustrated in figure 1. Let the projection matrices of images I and J be $\mathbf{P_I} = \mathbf{K}[\mathbf{I}|\mathbf{0}]$ and $\mathbf{P_J} = \mathbf{K}[\mathbf{R}|\mathbf{t}]$ with $\mathbf{K}$ the internal camera parameters and $[\mathbf{R}|\mathbf{t}]$ the pose increment of image J with respect to image I. If the image intensities represent linear X-Ray attenuation (as opposed to detector intensity), the value of a single pixel $I(\boldsymbol{x})$ at location $\boldsymbol{x}$ can be described as a line integral

$$I(\boldsymbol{x}) = \int_{\boldsymbol{y_1}}^{\boldsymbol{y_n}} \mu(\boldsymbol{r}) d\boldsymbol{r} \tag{1}$$

where $\mu : \mathcal{R}^3 \rightarrow \mathcal{R}$ is the volume of X-Ray attenuation, and $\boldsymbol{y_1}$, $\boldsymbol{y_n}$ the start and end position along the ray between source and detector position $\boldsymbol{x}$. As illustrated in figure 1 the attenuation values $\mu(\boldsymbol{r})$ along this ray also contribute

to pixel intensities in the neighboring projection image J. The affected pixels are located on a segment of the epipolar line of $\boldsymbol{x}$ in J. We can discretize and rewrite equation 1 as

$$I(\boldsymbol{x}) = \sum_{k=1}^{n} \mu(\boldsymbol{y_k}) + \mu_\epsilon \tag{2}$$

Here, $\boldsymbol{y_k}$ are sampling points along the projection ray going through $\boldsymbol{x}$ and μ_ϵ is the residual X-Ray attenuation not found within these sampling points. The attenuation values $\mu(\boldsymbol{y_k})$ contribute each with a particular fraction to the pixels $J(\boldsymbol{x_k})$:

$$\mu(\boldsymbol{y_k}) = w_k J(\boldsymbol{x_k}) \tag{3}$$

It is therefore valid to describe the relationship of a pixel $I(\boldsymbol{x})$ with the neigboring image as follows:

$$I(\boldsymbol{x}) = \sum_{k=1}^{n} w_k J(\boldsymbol{x_k}) + w_{n+1} \tag{4}$$

where $\boldsymbol{x_k}$ are a number of chosen discrete locations along the epipolar line in J. Essentially, the value of pixel $\boldsymbol{x}$ in I is composed of a linear combination of some of the X-Ray attenuation values its ray shares with the ones in the neighboring image's locations $\boldsymbol{x_k}$, plus a constant value for attenuation values outside the epipolar line segment.

If we restrict the locations of $\boldsymbol{x_k}$ such that the ray segment $|\boldsymbol{y_1} - \boldsymbol{y_n}|$ corresponds to the cylindrical core reconstruction area, the epipolar line segment in J should span from a few to maximum a dozen pixels (depending on image resolution, angle increment and reconstruction size). More precisely the start and end point of the epipolar line segment are given by

$$\begin{aligned} \boldsymbol{x_1} &= \mathbf{KRK}^{-1}\mathbf{x} + \frac{\mathbf{Kt}}{Z_{start}} \\ \boldsymbol{x_n} &= \mathbf{KRK}^{-1}\mathbf{x} + \frac{\mathbf{Kt}}{Z_{end}} \end{aligned} \tag{5}$$

where Z_{start} and Z_{end} are the start and end depths along the projection ray through $\boldsymbol{x}$ (for a derivation see [3] chapter 9). We linearly interpolate the samples in between as $\boldsymbol{x_k} = \boldsymbol{x_1} + \frac{k-1}{n-1}(\boldsymbol{x_n} - \boldsymbol{x_1})$.

The weights w_k basically describe at which depths in reconstruction space corresponding structures occur, and, consequently, in which horizontal direction and how fast they move in different "layers" in projection space. If they were known, one could assess how well the images satisfy equation 4.

2.2 Projection Similarity

Image similarity metrics which allow local variation of unknown parameters in the assumed intensity relationship, exhibit superior performance in certain X-Ray image registration problems. One particular example is local cross-correlation, where brightness and contrast may vary, since they are implicitly computed,

within the neighborhood of every image pixel [4]. It is also possible to implicitly recover unknown parameters of a linear combination of two image components within a similarity formulation [7].

We extend such an approach to create a similarity measure which assesses how well equation 4 locally holds. We use the following variables to specify the situation for a neighborhood Ω of size m surrounding a pixel $\boldsymbol{x}$

$$\boldsymbol{i}_x = \begin{pmatrix} i_1 \\ \vdots \\ i_m \end{pmatrix}; \mathbf{J}_x = \begin{pmatrix} j_{11} & \dots & j_{1n} & 1 \\ \vdots & & \vdots & \vdots \\ j_{m1} & \dots & j_{mn} & 1 \end{pmatrix}; \boldsymbol{w}_x = \begin{pmatrix} w_1 \\ \vdots \\ w_{n+1} \end{pmatrix} \tag{6}$$

where $\boldsymbol{i}_x$ is a linearization of the pixel intensities $\{I(\boldsymbol{x})|\boldsymbol{x} \in \Omega\}$ and row l of $\mathbf{J}_x$ contains n samples taken from the epipolar line corresponding to location $\boldsymbol{i}_l$ and a constant entry. We then interpret equation 4 as a linear regression problem

$$\boldsymbol{i}_x = \mathbf{J}_x \boldsymbol{w}_x \tag{7}$$

which we solve using ordinary least squares (OLS):

$$\boldsymbol{w}_x = (\mathbf{J}_x^T \mathbf{J}_x)^{-1} \mathbf{J}_x^T \boldsymbol{i}_x \tag{8}$$

The local similarity around pixel $\boldsymbol{x}$ then writes as

$$S_x(I, J) = 1 - \frac{|\boldsymbol{i}_x - \mathbf{J}_x \boldsymbol{w}_x|^2}{Var(\boldsymbol{i}_x)} \tag{9}$$

The global image similarity is composed weighted by the local variance, in order to suppress regions without structural appearance:

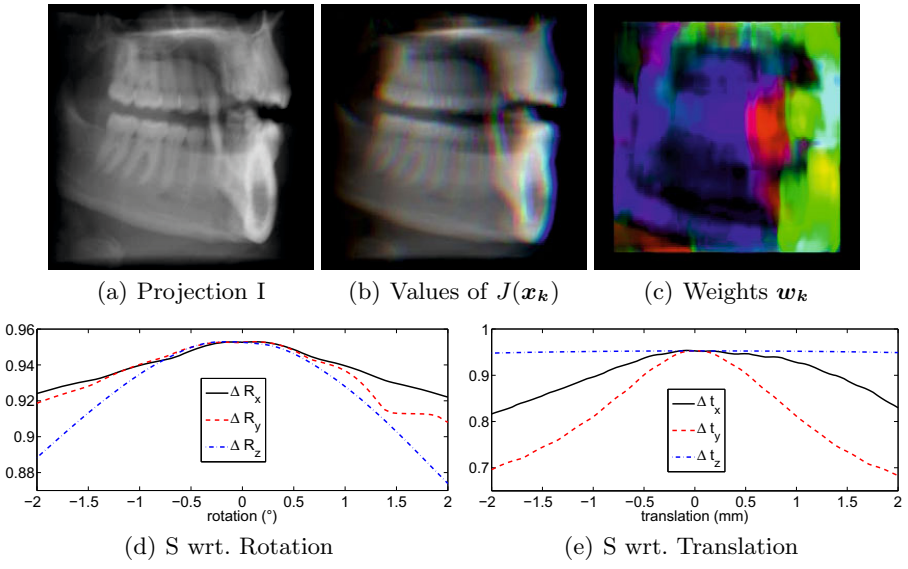
$$S(I, J) = \frac{\sum_{\boldsymbol{x}} \left(Var(\boldsymbol{i}_x) - |\boldsymbol{i}_x - \mathbf{J}_x \boldsymbol{w}_x|^2 \right)}{\sum_{\boldsymbol{x}} Var(\boldsymbol{i}_x)} = 1 - \frac{\sum_{\boldsymbol{x}} |\boldsymbol{i}_x - \mathbf{J}_x \boldsymbol{w}_x|^2}{\sum_{\boldsymbol{x}} Var(\boldsymbol{i}_x)} \tag{10}$$

Rather than computing vector $\boldsymbol{i}_x$ and matrix $\mathbf{J}_x$ for every pixel (and hence re-evaluating all neighborhood pixels repeatedly), we directly establish all required sums to solve equation 8 using an efficient recursive filtering approach.

2.3 Motion Detection

Depending on the parameters m and n which discretize the local X-Ray depth properties, the similarity measure value will slightly vary with changing image content even in absence of patient motion. On the other hand, it is dependant on the projection matrices $\mathbf{P_I}$ and $\mathbf{P_J}$. If we incorporate a rigid transformation matrix $\mathbf{T}(\boldsymbol{p}) = [\boldsymbol{\Delta}\mathbf{R}|\boldsymbol{\Delta}\mathbf{t}]$, parameterized over $\boldsymbol{p}$, the start and end point of the epipolar line segment of point $I(x)$ become

$$\begin{aligned} \mathbf{x_s} &= \mathbf{K}(\mathbf{R}\boldsymbol{\Delta}\mathbf{R})\mathbf{K}^{-1}\mathbf{x} + \frac{\mathbf{K}(\mathbf{R}\boldsymbol{\Delta}\mathbf{t} + \mathbf{t})}{Z_{start}} \\ \mathbf{x_e} &= \mathbf{K}(\mathbf{R}\boldsymbol{\Delta}\mathbf{R})\mathbf{K}^{-1}\mathbf{x} + \frac{\mathbf{K}(\mathbf{R}\boldsymbol{\Delta}\mathbf{t} + \mathbf{t})}{Z_{end}} \end{aligned} \tag{11}$$



(a) Projection I (b) Values of $J(\boldsymbol{x_k})$ (c) Weights $\boldsymbol{w_k}$

(d) S wrt. Rotation (e) S wrt. Translation

Fig. 2. Illustration of similarity measure properties from a pair of DRRs (please refer to the colored electronic version of this manuscript)

Even though it is just defined on the 2D projection images, the measure S will change with respect to all 3D transformation parameters, because $\mathbf{T}(\boldsymbol{p})$ directly affects the pixel locations $\boldsymbol{x_k}$ in equation 4. If there is no patient motion the similarity measure should have a maximum at $\mathbf{T}(\mathbf{0}) = \mathbf{I}$, i.e. $\frac{\partial S}{\partial \boldsymbol{p}}|_{\boldsymbol{p}=\mathbf{0}}$ should be zero. Due to the local least-squares solutions inherent to the measure, analytic gradients are difficult to derive. We therefore use a numeric difference operator in order to assess if $S(I, J, \boldsymbol{p})$ is at a maximum. More precisely, we sum over all forward and backward differences which are positive:

$$S' = \sum_{i=1}^{12} \frac{1}{2} \left(\operatorname{sgn}(\Delta S_i) + 1\right) \Delta S_i \quad \text{with } \Delta S_i = S(\Delta_i) - S(\mathbf{0}) \tag{12}$$

3 Evaluation

3.1 Analysis of the Similarity Value

We use a high-fidelity dental cone beam reconstruction with 512^3 voxels and compute digitally reconstructed radiographs (DRRs) from it. Their quality is insignificantly lower than the original projection images, however they enable us to simulate arbitrary motion of the patient's head in known conditions. Figure 2 visualizes the similarity measure. A parameter selection of $n = 3$ and $m = 21^2$ was used. The three values $J(\boldsymbol{x}_i)$, as well as the weights w_i, are encoded in the red, green and blue color channels of figure 2(b). The distribution

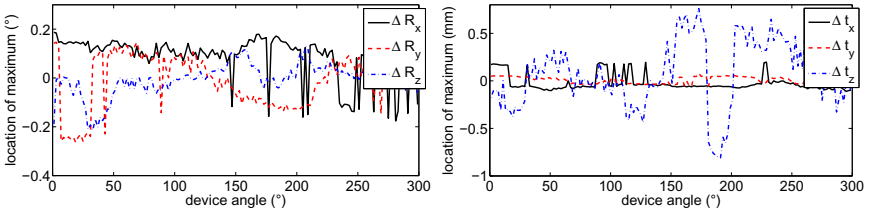


Fig. 3. Location of local maxima for the parameters of $\mathbf{\Delta R}$ and $\mathbf{\Delta t}$

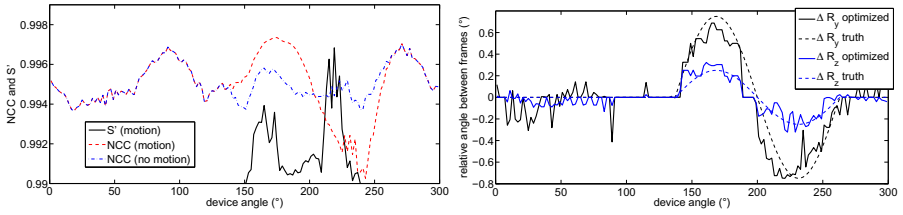
along the epipolar line can nicely be seen by the horizontal split into color components, in particular at vertical structures. The computed weights w_i in figure 2(c) essentially represent a rough depth reconstruction from the two projections. Structures in the front of the volume which are moving to the right in the projections are blue. Structures in the center of the volume or tangential to the rotation are green; structures in the back (moving to the left) are red. The similarity measure S in equation 10 was evaluated with respect to all parameters in $\mathbf{\Delta R}$ and $\mathbf{\Delta t}$, figure 2(d) and (e) depict the corresponding curves for the image pair shown. This evaluation was then repeated for all projection pairs of a simulated cone-beam acquisition without any patient motion (300° rotation, 2° angle increment). Figure 3 shows the location of the similarity measure maximum for all frames (it should be zero). The mean deviation and variance (σ) are:

	$\mathbf{\Delta R}_x$	$\mathbf{\Delta R}_y$	$\mathbf{\Delta R}_z$	$\mathbf{\Delta t}_x$	$\mathbf{\Delta t}_y$	$\mathbf{\Delta t}_z$
mean	0.0892	-0.0024	-0.0040	-0.0214	0.0060	0.0754
σ	0.0870	0.1173	0.0655	0.0879	0.0351	0.3542

All parameters vary within a very small scale. The variation of the translation $\mathbf{\Delta t}_z$ towards the X-Ray source is highest, which is expected since it causes very little change in the projections (compare fig. 2(e)).

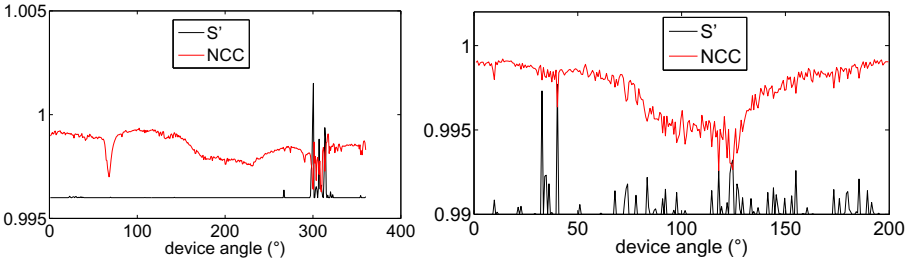
3.2 Recovering Motion

A cone-beam acquisition was simulated which includes a gentle nod of the patient's head with an amplitude of 10° and 7° in the rotational parameters $\mathbf{\Delta R}_y$ and $\mathbf{\Delta R}_z$. Here, the first parameter corresponds to a relative change of the cone-beam rotation speed, therefore the combined rotation slows down first. In this case, a standard similarity measure will yield higher similarity, as can be seen in figure 4(a) for the normalized cross-correlation (NCC) similarity. Our detection S' accurately reflects the amount of motion (the plot shows $S' + 0.99$ to fit on the same scale). S' is zero throughout the sequence without motion (hence not included in the plot). Figure 4(b) shows the derivative of our simulated motion (i.e. the motion per frame pair) along with the result of a full non-linear optimization over S', individually for each image pair. Even though the incremental patient motion between frames has a maximum amplitude of only $\approx 0.75°$, our method can approximately recover it with an average deviation of $0.093 \pm 0.122°$ in $\mathbf{\Delta R}_y$ and $0.046 \pm 0.059°$ in $\mathbf{\Delta R}_z$.



(a) motion detection (b) recovered motion parameters

Fig. 4. Result of the motion detection experiment



(a) Patient 1 (b) Patient 2

Fig. 5. Motion detection on clinical projection data from two patients

3.3 Parameter Selection

The parameters of our system to be defined up-front, namely the local neighborhood size m and the number of motion layers n, have to be chosen in order to optimize the similarity robustness for a particular clinical application or anatomy. While a systematic evaluation is beyond the scope of this manuscript, we have generally obtained good results for $m = 21^2 \dots 31^2$ and $n = 3 \dots 7$ on various anatomy, given a projection image size of 512^2 pixels. Those parameters are also intuitively defined; in dental applications $n = 3$ is sufficient since the structures are dominated by two layers of teeth moving in opposite direction (see figure 2). The maximum reasonable value of n is where $|\boldsymbol{x}_{i+1} - \boldsymbol{x}_i|$ is roughly the size of an individual image pixel (see equation 11).

3.4 Real Projection Data

We have computed the motion detection measure S' on the original projection images of two orthodontic cone-beam acquisitions. Patient 1 exerts a strong head movement towards the end of the sequence. Patient 2 is trembling several times during the entire acquisition. Figure 5 depicts the detection value S' for all frame pairs, and the NCC measure for comparison. The value S' only deviates from zero where patient motion is clearly visible.

4 Conclusion

As opposed to disparity computation from stereo cameras, a pair of X-Ray images yields correspondence at multiple depths. We were able to approximately recover this information as a number of discrete weights, through a local least-squares estimation for every pixel's neighborhood. This in turn allowed us to derive an optimal similarity formulation in projection space, which is sensitive to 3D object changes beyond the expected relation of the projections themselves. We have shown on dental cone-beam data that this measure accurately detects global rigid patient motion. Besides, we could estimate the intra-frame motion parameters with an accuracy of $< 0.1°$ and $< 0.4mm$, respectively.

We believe that this formulation can help future reconstruction algorithms to better take patient motion into account. Equation 12 can detect motion right during the acquisition and suggest if the reconstruction needs to consider it in the first place. While our actual motion estimates from X-Ray image pairs alone would yield drift applied over a sequence, they can bootstrap more powerful iterative methods which simultaneously estimate a volume and motion parameters. The computed depth weights w_i are a by-product of our formulation, and could be exploited to accelerate the back-projection step during a reconstruction. Extensions to use more than two projections, as well as non-linear motion models, are straightforward.

References

1. Ens, S., Ulrici, J., Hell, E., Buzug, T.: Automatic detection of patient motion in cone-beam computed tomography. In: 2010 IEEE International Symposium on Biomedical Imaging, pp. 1257–1260 (2010)
2. Hansis, E., Schomberg, H., Erhard, K., Dössel, O., Grass, M.: Four-dimensional cardiac reconstruction from rotational x-ray sequences - first results for 4D coronary angiography. In: SPIE Medical Imaging 2009 Conference, vol. 7258 (2009)
3. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision, 2nd edn. Cambridge University Press, Cambridge (2004)
4. Khamene, A., Bloch, P., Wein, W., Svatos, M., Sauer, F.: Automatic registration of portal images and volumetric CT for patient positioning in radiation therapy. Medical Image Analysis, 96–112 (February 2006)
5. Rohkohl, C., Lauritsch, G., Biller, L., Hornegger, J.: ECG-gated interventional cardiac reconstruction for non-periodic motion. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6361, pp. 151–158. Springer, Heidelberg (2010)
6. Schretter, C., Pilatus, F., Rose, G., Aach, T., Bertram, M.: Optical flow motion estimation for approximate motion compensation in cone-beam CT. In: Proceedings of the 10th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine, pp. 327–330 (2009)
7. Wein, W., Brunke, S., Khamene, A., Callstrom, M., Navab, N.: Automatic CT-ultrasound registration for diagnostic imaging and image-guided intervention. Medical Image Analysis 12, 577–585 (2008)

Multi-feature Statistical Nonrigid Registration Using High-Dimensional Generalized Information Measures

Sameh Hamrouni[1], Nicolas Rougon[1], and Françoise Prêteux[2]

[1] ARTEMIS Department; CNRS UMR 8145 - TELECOM SudParis, Evry, France
{Sameh.Hamrouni,Nicolas.Rougon}@it-sudparis.eu
[2] Direction des Recherches - Mines ParisTech, Paris, France

Abstract. Nonrigid image registration methods based on the optimization of information-theoretic measures provide versatile solutions for robustly aligning mono-modal data with nonlinear variations and multi-modal data in radiology. Whereas mutual information and its variations arise as a first choice, generalized information measures offer relevant alternatives in specific clinical contexts. Their usual application setting is the alignement of image pairs by statistically matching scalar random variables (generally, greylevel distributions), handled via their probability densities. In this paper, we address the issue of estimating and optimizing generalized information measures over high-dimensional state spaces to derive *multi-feature* statistical nonrigid registration models. Specifically, we introduce novel consistent and asymptotically unbiaised k nearest neighbors estimators of α-informations, and study their variational optimization over finite and infinite dimensional smooth transform spaces. The resulting theoretical framework provides a well-posed and computationally efficient alternative to entropic graph techniques. Its performances are assessed on two cardiological applications: measuring myocardial deformations in tagged MRI, and compensating cardio-thoracic motions in perfusion MRI.

Keywords: Multi-feature nonrigid registration, groupwise nonrigid registration, kNN entropy estimators, high-dimensional α-information, cardiac tagged MRI, cardiac perfusion MRI.

1 Introduction

Nonrigid image registration methods based on the optimization of information measures provide versatile solutions for robustly aligning mono-modal data with nonlinear variations and multi-modal (inhomogeneous or incommensurable) data in medical radiology. Whereas mutual information and its variations arise as a first choice [8], generalized information measures, such as f-informations [11,12] and Jensen f-divergences [9], have been shown to be relevant alternatives in specific clinical contexts.

Their usual application setting is the alignement of image pairs by statistically matching luminance distributions, handled using marginal and joint probability densities estimated via kernel techniques. Though efficient for joint densities exhibiting well-separated clusters or reducible to simple mixtures, these approaches reach their limits for nonlinear mixtures, such as generated by textured objects over complex backgrounds. In this

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 524–531, 2011.

case, pixelwise luminance appears to be a too coarse feature for allowing unambiguous local statistical decisions, resulting into inaccurate or inconsistent matches, especially problematic in atlas propagation or motion estimation frameworks.

An intuitive idea to overcome this limitation is to combine multiple higher-dimensional image features (which may convey spatial information) within a vector random variable (RV). Its implementation, however, is faced with the curse of dimensionality: the number of samples required to consistently estimate a probability density increases exponentially with the state space dimension. In particular, kernel density estimators fail in high dimensions due to the fixed size of sample sets, which precludes estimating and optimizing information measures using plug-in techniques. Groupwise registration of image series, *i.e.* their simultaneous alignment onto a common reference, comes up against a similar obstacle, linked to handling spatio-temporal luminance distributions.

This theorerical issue can be bypassed by computing information measures directly from data using *geometric entropy estimators*. The latter include (i) entropic graphs, applicable to the broader class of α-informations, whose relevance for image registration is today well-established [9,13]; and (ii) more recent kth-nearest neighbor (kNN) entropy estimators [4,7], whose application to image science has been until today confined to the Shannon framework [3,5,6].

In this paper, we introduce novel, consistent and asymptotically unbiased kNN estimators of α-informations, and study their variational optimization over finite- and infinite-dimensional spaces of smooth spatial mappings. The resulting theoretical framework provides a well-posed and computationally efficient alternative to entropic graph techniques for multi-feature nonrigid image registration. Its performances are assessed in cardiac MRI for two clinical challenges: measuring myocardial deformations in tagged MRI, and compensating cardio-thoracic motions in perfusion MRI.

2 Multi-feature Information-Theoretic Registration

2.1 Problem Statement and Modeling

In order to build a unified formalism including pairwise and groupwise approaches, we deal with the general issue of registering synchronized image sequences. A sequence of 2D greylevel images defined over a domain $\Omega \subset \mathbb{R}^2$ is modeled as a mapping $I : \Omega \times [1..d] \to \Lambda$ with values in an interval $\Lambda \subset \mathbb{Z}$. Given a source sequence I_M comprising d unregistered images, and a reference sequence I_T with the same length, let x_τ denote the spatial coordinates[1] in frame $I_M(\cdot, \tau)$ and let $x = [x_1...x_d] \in \Omega^d$. We seek for a d-dimensional vector of smooth spatial transforms $\phi(x) = [\phi_1(x_1)...\phi_d(x_d)]$ in some space $\mathcal{T}$ such that ϕ_τ ($\tau \in [1..d]$) maps the source frame $I_M(\cdot, \tau)$ onto its analog $I_T(\cdot, \tau)$ in the reference sequence. The case $d = 1$ amounts to pairwise registration. Groupwise registration corresponds to $d > 1$, I_T being either a registered sequence or the concatenation of d instances of the same reference image.

Local correspondences are established via a dense set M (resp. T) of spatio-temporal features extracted from I_M (I_T) with values in a D-dimensional space $\mathcal{F}$, which can comprise both radiometric and spatial information. Features are integrated in vectors

[1] Similarly, the τth component of a vector r is hereafter denoted by r_τ.

$M(x) = [M(x_1, 1)...M(x_d, d)] \in \mathcal{F}$ (resp. $T(x)$)[2]. To robustly account for sensor-, patient- and pathology-related variability, we adopt a statistical framework, and model M and T as random variables (RV) over the state space $\mathcal{F}$, with respective marginal densities p^M and p^T, joint density $p^{M,T}$ and copula density $\pi^{M,T} = \frac{p^{M,T}}{p^M p^T}$.

Registration is then tackled as an optimal statistical matching problem between feature distributions over the transform space $\mathcal{T}$: $\phi^* = \operatorname{argmin}_{\phi \in \mathcal{T}} \mathcal{C}(M, T, \phi)$. The matching criterion $\mathcal{C}(M, T, \phi)$ is expressed as the sum of a discrepancy term $\mathcal{S}(M^\phi, T)$ between the reference RV T and the RV M^ϕ computed from the warped sequence $I_M^\phi = I_M \circ \phi^{-1}$, shortly denoted by M; and a regularization term $\mathcal{R}(\phi)$ ensuring smooth solutions. In this paper, we study the α-information family, comprising the pseudo-additive Havrda-Charvát information $I_\alpha(M, T) = \frac{1}{\alpha - 1}\left(J_\alpha(M, T) - 1\right)$ and the additive Renyi information $I_\alpha^*(M, T) = \frac{1}{\alpha - 1} \log J_\alpha(M, T)$, where:

$$J_\alpha(M, T) = \int_{\mathcal{F}^2} p^{M,T}(m, t) \left(\pi^{M,T}(m, t)\right)^{\alpha - 1} dm dt = \mathbb{E}_{M,T}\left[\left(\pi^{M,T}\right)^{\alpha - 1}\right]$$

These measures, defined for $\alpha \in \mathbb{R}^+ \backslash \{0, 1\}$, generalize mutual information which is obtained in the limit $\alpha \to 1$, yielding discrepancy functionals $\mathcal{S} = -I_\alpha$ and $\mathcal{S} = -I_\alpha^*$.

2.2 kNN Estimators of α-Informations

Given a q-dimensional RV Y, let $Y(\Omega^d)$ be a set of i.i.d. samples $Y(x)$ of Y indexed by $x \in \Omega^d$. kNN estimation techniques aim at characterizing Y from the statistics of q-dimensional balls $\mathcal{B}_k^Y(x) \subset \mathbb{R}^q$, centered at sample $Y(x)$ and containing its k nearest neighbors in $Y(\Omega^d) \backslash \{Y(x)\}$. The resulting estimators are functions of the radius $\rho_k^Y(x)$ of $\mathcal{B}_k^Y(x)$, equal to the Euclidean distance from $Y(x)$ to its kth nearest neighbor in this set. An early example is the Loftsgaarden-Quesenberry density estimator: $p_{\mathrm{LQ}}^Y(Y(x)) = \frac{k}{|\Omega|\, V_q \left(\rho_k^Y(x)\right)^q}$. Though notoriously biased in high dimension, the latter leads however to a consistent and asymptotically unbiased kNN estimator of Shannon entropy [4], from which kNN estimators of Kullback-Leibler divergence [4,3], and mutual information and its variations [5,6] have been derived.

This result has been recently extented to α-entropies [7], based on a consistent and asymptotically unbiased kNN estimator of the integral $J_\alpha(Y) = \int_{\mathbb{R}^q} \left(p^Y(y)\right)^\alpha dy$:

$$J_\alpha^{\mathrm{knn}}(Y) = \frac{1}{|\Omega|} \sum_{x \in \Omega^d} \left((|\Omega| - 1) C_k(\alpha) V_q \left[\rho_k^Y(x)\right]^q \right)^{1 - \alpha} \tag{1}$$

where $C_k(\alpha) = \left[\frac{\Gamma(k)}{\Gamma(k + 1 - \alpha)}\right]^{\frac{1}{1-\alpha}}$ and $V_q = \frac{\pi^{q/2}}{\Gamma(q/2 + 1)}$ is the volume of the unit ball of $\mathbb{R}^q$. $J_\alpha^{\mathrm{knn}}(Y)$ can be interpreted as a plug-in estimator, built from an Ahmad-Lin estimator [1] $J_\alpha^{\mathrm{AL}}(Y) = \frac{1}{|\Omega|} \sum_{x \in \Omega^d} \left(p^Y\left(Y(x)\right)\right)^{\alpha - 1}$ of $J_\alpha(Y) = \mathbb{E}_Y\left[\left(p^Y\right)^{\alpha - 1}\right]$ using the following kNN density estimator:

$$p_{\mathrm{knn}}^Y(Y(x)) = \frac{1}{(|\Omega| - 1) C_k(\alpha) V_q \left(\rho_k^Y(x)\right)^q}$$

[2] Here, $M(x_\tau, \tau)$ is a D_1-dimensional feature vector extracted from $I_M(\cdot, \tau)$ with $D = dD_1$.

Along the same lines, we derive high-dimensional kNN estimators of α-informations by introducing a novel consistent and asymptotically unbiased estimator of the integral $J_\alpha(M,T)$, defined as:

$$J_\alpha^{\mathrm{knn}}(M,T) \;=\; \frac{1}{|\Omega|}\sum_{x\in\Omega^d}\left((|\Omega|-1)C_k(\alpha)\frac{V_D^2}{V_{2D}}\left[\frac{\rho_k^M(x)\rho_k^T(x)}{(\rho_k^{M,T}(x))^2}\right]^D\right)^{\alpha-1} \tag{2}$$

Building this estimator, which is achieved similarly to the entropy estimator $J_\alpha^{\mathrm{knn}}(Y)$, is formally equivalent to plugging the kNN copula density estimator $\pi_{\mathrm{knn}}^{M,T} = \frac{p_{\mathrm{knn}}^{M,T}}{p_{\mathrm{knn}}^{M} p_{\mathrm{knn}}^{T}}$ of $\pi^{M,T}$ in the following Ahmad-Lin estimator of $J_\alpha(M,T)$:

$$J_\alpha^{\mathrm{AL}}(M,T) \;=\; \frac{1}{|\Omega|}\sum_{x\in\Omega^d}\left(\pi^{M,T}(x)\right)^{\alpha-1}$$

where $\pi^{M,T}(x) = \pi^{M,T}(M(x),T(x))$. From (2), we obtain consistent estimators $I_\alpha^{\mathrm{knn}}(M,T) = \frac{1}{\alpha-1}\left(J_\alpha^{\mathrm{knn}}(M,T)-1\right)$ and $I_\alpha^{*\mathrm{knn}}(M,T) = \frac{1}{\alpha-1}\log J_\alpha^{\mathrm{knn}}(M,T)$ of Havrda-Charvát and Renyi information, respectively. In addition, $I_\alpha^{\mathrm{knn}}(M,T)$ is asymptotically unbiased.

2.3 Variational Optimization

Searching for the optimal transform ϕ^* using gradient descent requires a closed-form expression of the first variation of α-informations over the transform space $\mathcal{T}$. Since the kNN estimator $J_\alpha^{\mathrm{knn}}(M,T)$ is not differentiable, a two-step plug-in approach is used. Starting from consistent Ahmad-Lin estimators $I_\alpha^{\mathrm{AL}}(M,T) = \frac{1}{\alpha-1}\left(J_\alpha^{\mathrm{AL}}(M,T)-1\right)$ and $I_\alpha^{*\mathrm{AL}}(M,T) = \frac{1}{\alpha-1}\log J_\alpha^{\mathrm{AL}}(M,T)$, we first classically compute a closed-form expression of their variational derivatives using differentiable kernel density estimates and a Mean-Shift approximation of the ratio $\nabla p/p$ for the densities p^M and $p^{M,T}$. We then switch to the kNN framework by considering uniform kernels over the balls $\mathcal{B}_k^M$ and $\mathcal{B}_k^{M,T}$ (their derivatives being indicators over the spheres $\mathcal{S}_k^M$ and $\mathcal{S}_k^{M,T}$). This strategy yields consistent kNN estimators of the variational derivative of α-informations.

For infinite dimensional transform spaces, consisting of smooth mappings $\phi_\tau = \mathbb{I}\mathrm{d} + u_\tau$ described by displacement fields u_τ to be estimated at all $x\in\Omega^d$, we get:

$$\partial_{u_\tau} J_\alpha^{\mathrm{knn}}(x) \;=\; \mathcal{V}(x)\,\mathcal{D}_\tau^M(x)$$

where $\mathcal{V}(x) = c_3\left[K_k^{M,T}(x)\left(\mathcal{L}_{k,\tau}^M(x)-\mathcal{L}_{k,\tau}^{M,T}(x)\right)+c_2\left(L_{k,\tau}^M(x)-L_{k,\tau}^{M,T}(x)\right)\right]$ and $\mathcal{D}_\tau^M(x) = (\partial_{x_\tau}M(x))^T$. Letting $\mathcal{N}_k^{M,T}(x) = \{\, y\in\Omega^d \,|\, x\in\mathcal{S}_k^{M,T}(y)\,\}$, we have:

$$\mathcal{L}_{k,\tau}^{M,T}(x) = \frac{2d+2}{\rho_k^{M,T}(x)^2}\sum_{y\in\mathcal{B}_k^{M,T}(x)} r^M(x,y) \qquad L_{k,\tau}^{M,T}(x) = \sum_{y\in\mathcal{N}_k^{M,T}(x)}\frac{r^M(x,y)}{|r^M(x,y)|}\left(\pi_{\mathrm{knn}}^{M,T}(y)\right)^{\alpha-1}$$

$$\mathcal{L}_{k,\tau}^{M}(x) = \frac{d+2}{\rho_k^{M}(x)^2}\sum_{y\in\mathcal{B}_k^{M}(x)} r^M(x,y) \qquad L_{k,\tau}^{M}(x) = \sum_{y\in\mathcal{N}_k^{M}(x)}\frac{r^M(x,y)}{\rho_k^M(y)}\left(\pi_{\mathrm{knn}}^{M,T}(y)\right)^{\alpha-1}$$

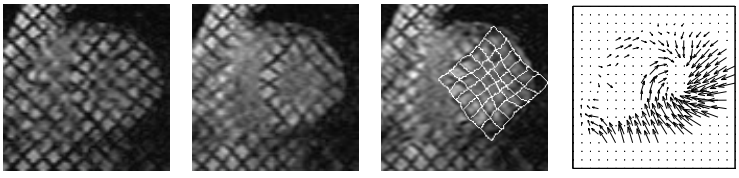


Fig. 1. Estimating myocardial deformations in tagged MRI using multi-feature Havrda-Charvát information-minimizing nonrigid registration

Finally: $K_k^{M,T}(x) = \frac{1}{k}\big(\pi_{\mathrm{knn}}^{M,T}(x)\big)^{\alpha-1}$, $c_2 = C_k(\alpha)\,\frac{|\Omega|-1}{|\Omega|}$ and $c_3 = \frac{\alpha-1}{|\Omega|}$.

The optimal transforms ϕ_τ^* are iteratively estimated at every pixel $x \in \Omega^d$ with the explicit scheme: $u_\tau^{t+1}(x) = u_\tau^t(x) + \delta_t\,\partial_{u_\tau} {I_\alpha}^{\mathrm{knn}}(x)$ and accordingly for ${I_\alpha}^{*\mathrm{knn}}$.

For finite dimensional transform spaces, consisting of smooth mappings $\phi_\tau(x_\tau) = B(x_\tau)\Theta_\tau$ described by a parameter $\Theta_\tau \in \mathbb{R}^p$ in a local basis encoded by the $(2 \times p)$ matrix $B(x_\tau)$, we get:

$$\partial_{\Theta_\tau} J_\alpha^{\mathrm{knn}} = \sum_{x \in \Omega^d} \mathcal{V}(x)\,\mathcal{D}_{B,\tau}^M(x)$$

where $\mathcal{D}_{B,\tau}^M(x) = \left(\partial_{x_\tau} M(x)\right)^T B(x_\tau)$. The optimal parameters Θ_τ^* are computed using the update equation: $\Theta_\tau^{t+1} = \Theta_\tau^t + \delta_t\,\partial_{\Theta_\tau} {I_\alpha}^{\mathrm{knn}}(\Theta^t)$ and accordingly for ${I_\alpha}^{*\mathrm{knn}}$.

2.4 Implementation Details

The numerical complexity of the optimization scheme, dictated by the kNN search algorithm, is reduced by using an approximate nearest neighbor technique [2] with *linear* complexity and memory usage w.r.t. the number of samples $|\Omega|$ and feature dimension D. Obviously, complexity increases with k. In practise, this parameter is tuned empirically by searching for a trade-off between registration accuracy and processing time. For the cardiac MRI applications hereafter, setting $k = 15$ yielded satisfying results.

3 Application to Cardiac MRI

3.1 Estimating Myocardial Motion in Tagged MRI Using Pairwise Registration

The proposed model has been applied to the estimation of myocardial deformations from tagged MRI exams by sequentially performing frame-to-frame registration [12]. We used a local feature vector with normalized components, comprising the greylevel, and the eigenvalues and eigenvectors of the structure tensor $\partial_{x_\tau} I(x)^T \partial_{x_\tau} I(x) \star G_\sigma$ of the image I, computed at a Gaussian scale σ equal to the half-width of the tagging pattern ($D = 5$). We studied the impact on registration accuracy of using a directional local contrast descriptor by comparing performances with a similar registration model exploiting solely the greylevel information [12]. To be consistent with the later work, we considered an infinite dimensional (non-parametric) transform space, and a registration

Table 1. MSE (in pixels) between estimated and ground-truth motion magnitude for simulated tagged MRI datasets using greylevel ($D = 1$) and greylevel + structure tensor ($D = 5$) features

D	Base	Median	Apex
1	0.73 ± 0.31	0.88 ± 0.51	1.53 ± 0.75
5	0.35 ± 0.22	0.42 ± 0.34	0.65 ± 0.59

Table 2. MSE and Pearson correlation coefficient (R^2) between estimated and ground-truth parameters as a function of maximal displacement δ (in pixels) for simulated p-MRI datasets at median level with random FFD misalignment

	δ	1	3	5	7	10
MSE	NMI	0.52	0.76	0.63	0.83	1.5
	I^*_α	0.41	0.49	0.57	0.62	0.80
R^2	NMI	0.94	0.91	0.93	0.90	0.83
	I^*_α	0.95	0.94	0.93	0.93	0.91

Table 3. Dice similarity indices between reference and registration-induced segmentations of the left ventricle (Lv), right ventricle (Rv) and myocardium (Myo)

	Lv	Rv	Myo
native	0.89	0.003	0.20
registered using NMI	0.98	0.95	0.88
registered using I^*_α	0.98	0.87	0.93

criterion combining Havrda-Charvát information ($\alpha = 1.2$) and the Nagel-Enkelmann oriented smoothness stabilizer, and performed experiments on the SPAMM dataset used in [12]. Qualitative assessment of the estimated displacement fields by an expert cardiologist indicate systematic improvement of motion estimates (i) on endo- and epicardial boundaries, where the tagging signal suffers from poor localization and contrast fading artifacts, and (ii) over the whole myocardium at end of sequence (Figure 1). This was confirmed on synthetic tagged MRI sequences generated from natural exams by warping their first frame with the statistical myocardial motion atlas described in [10]. Comparing the MSE between ground-truth and estimated motion over the dataset shows that accuracy is improved by a multi-feature registration model (Table 1).

3.2 Groupwise Registration of Perfusion MRI Exams

The proposed model has been also applied to the compensation of cardio-thoracic motions in partial breath-hold and free-breathing contrast-enhanced cardiac perfusion MRI (p-MRI) exams. Here, the challenge lies in aligning structures exhibiting highly nonlinear contrast variations and complex deformations. We adopt a groupwise registration strategy involving a motion-free reference p-MRI exam and using pixel-wise contrast enhancement curves $M(x) = [I_M(x_1, 1)...I_M(x_d, d)]$ (resp. $T(x)$) as local features ($D = d \in [25..35]$). Using spatio-temporal features induces an alignment process driven by the statistical properties of the *whole exam*, with the expected benefit of an improved alignment overall consistency compared to classical pairwise schemes based

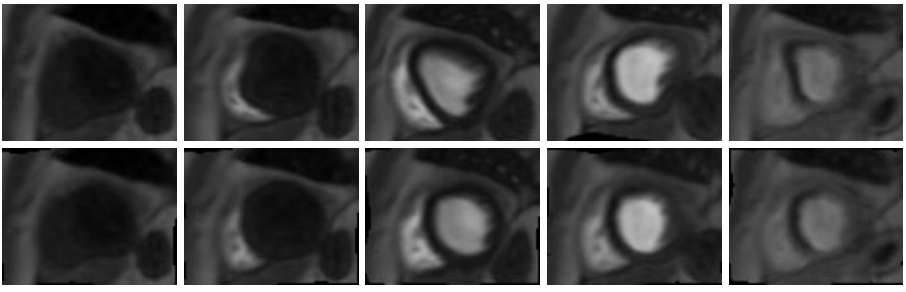


Fig. 2. Evaluation of registration accuracy. *Top:* Simulated unregistered exam ($\delta = 7$) at median level ($d = 34$ frames). *Bottom:* registration over a reference motion-free exam I_T using I^*_α.

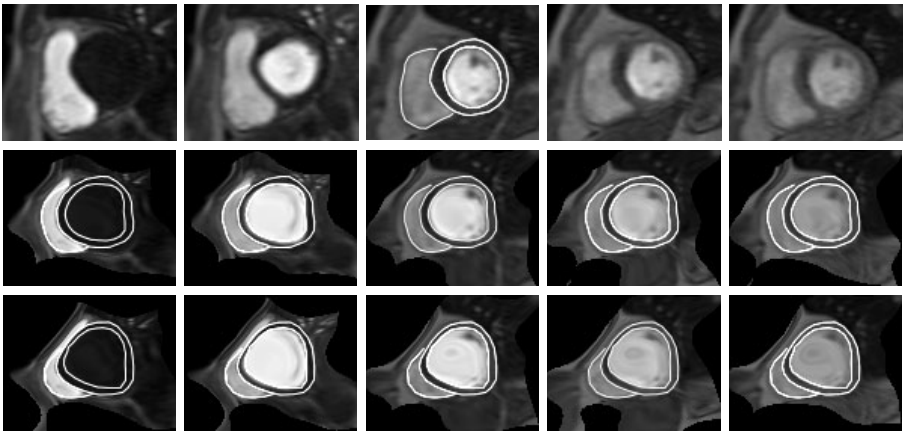


Fig. 3. Evaluation of registration robustness. *Top*: unregistered synthetic exam at median level created by concatenating $3 + 3$ frames from 2 different patient exams. *Middle (resp. bottom):* registration using NMI (resp. I^*_α). *Overlay*: registration-induced segmentation derived from a supervised delineation of the heart in a reference frame (*top*) replicated on all images. Warping artefacts over thoracic regions originate from using a synthetic template with uniform background.

on a reference frame with maximal myocardial contrast. We use a matching criterion combining Renyi information ($\alpha = 1.2$) and a thin-plate spline stabilizer, which is optimized over a Free-Form Deformation (FFD) basis parameterized by 8×8 control grids. Registration results were systematically compared with those delivered by a similar model [6] using a high-dimensional kNN estimator of normalized mutual information (NMI) instead of Renyi information. To this end, the dataset and experimental validation protocol described in [6] were used. Results on synthetic sequences, generated by applying to the reference exam random FFD transforms with control point displacements uniformly distributed in the $[-\delta, \delta]$ pixels range (Figure 2), show that Renyi information leads to improving registration accuracy (Table 2). Similar conclusions were reached for registration robustness (Table 3), which was assessed on composite datasets synthetized by intermixing frames from different patient exams at the same slice level

(Figure 3). Processing $d = 35$ frames at 3 slice levels (*i.e.* base, medium, apex) requires approximately 13 minutes on a standard 2.4GHz dual core PC with 2Gb RAM.

4 Conclusion

We have presented novel kNN estimators of high-dimensional α-informations, and derived closed-form expressions of their first variation over finite- and infinite-dimensional spaces of smooth spatial tranforms. This theoretical framework enables variational multifeature nonrigid registration based on a non-Shannon information-theoretic model, and provides a computationally efficient alternative to entropic graph techniques. The formalism, developed for the alignement of 2D image sequences, encompasses both pairwise and groupwise registration, and is obviously readily valid in 3D. Future work is currently directed towards increasing computational performances by using more efficient gradient descent schemes and boosting kNN search using GPU implementations.

References

1. Ahmad, I., Lin, P.: A nonparametric estimation of the entropy for absolutely continuous distributions. IEEE Transactions on Information Theory 22(3), 372–375 (1976)
2. Arya, S., Mount, D.M., Netanyahu, N.S., Silverman, R., Wu, A.: An optimal algorithm for approximate nearest neighbor searching. Journal of the ACM 45(6), 891–923 (1998)
3. Boltz, S., Debreuve, E., Barlaud, M.: High-dimensional statistical measure for region-of-interest tracking. IEEE Transactions on Image Processing 18(6), 1266–1283 (2009)
4. Goria, M., Leonenko, N., Mergel, V., Novi Inverardi, P.L.: A new class of random vector entropy estimators and its applications in testing statistical hypotheses. Journal of Nonparametric Statistics 17(3), 277–297 (2005)
5. Hamrouni, S., Rougon, N., Prêteux, F.: Multi-feature information-theoretic image registration: application to groupwise registration of perfusion MRI exams. In: Proceedings IEEE International Symposium on Biomedical Imaging: Fron Nano to Macro, Chicago, IL (2011)
6. Hamrouni, S., Rougon, N., Prêteux, F.: Groupwise registration of cardiac perfusion MRI sequences using normalized mutual information in high dimension. In: Proceedings SPIE Medical Imaging 2011 - Image Processing, Orlando, FL, vol. 7962 (2011)
7. Leonenko, N., Pronzato, L., Savani, V.: A class of Rényi information estimators for multidimensional densities. Annals of Statistics 36(5), 2153–2182 (2008)
8. Maes, F., Vandermeulen, D., Suetens, P.: Medical image registration using mutual information. Proceedings of the IEEE 91(10), 1699–1722 (2003)
9. Neemuchwala, H.F., Hero, A.O.: Entropic graphs for registration. In: Multi-sensor Image Fusion and its Applications. Marcel Dekker, New York (2004)
10. Petitjean, C., Rougon, N., Prêteux, F.: Building and using a statistical 3D motion atlas for analyzing myocardial contraction in MRI. In: Proceedings SPIE Medical Imaging 2004 - Image Processing, San Diego, CA, vol. 5370, pp. 253–264 (2004)
11. Pluim, J.P.W., Maintz, J.B.A., Viergever, M.A.: f-information measures in medical image registration. IEEE Transactions on Medical Imaging 23(12), 1508–1516 (2004)
12. Rougon, N., Petitjean, C., Prêteux, F., Cluzel, P., Grenier, P.: A non-rigid registration approach for quantifying myocardial contraction in tagged MRI using generalized information measures. Medical Image Analysis 9(4), 353–375 (2005)
13. Staring, M., van der Heide, U.A., Klein, S., Viergever, M.A., Pluim, J.P.W.: Registration of cervical MRI using multifeature mutual information. IEEE Transactions on Medical Imaging 28(9), 1412–1421 (2009)

Estimating the 4D Respiratory Lung Motion by Spatiotemporal Registration and Building Super-Resolution Image

Guorong Wu[1], Qian Wang[1], Jun Lian[2], and Dinggang Shen[1]

[1] Department of Radiology and BRIC, University of North Carolina at Chapel Hill
{grwu,dgshen}@med.unc.edu, qiangwang@email.unc.edu
[2] Department of Radiation Physics, University of North Carolina at Chapel Hill
jun_lian@med.unc.edu

Abstract. The estimation of lung motion in 4D-CT with respect to the respiratory phase becomes more and more important for radiation therapy of lung cancer. Modern CT scanner can only scan a limited region of body at each couch table position. Thus, motion artifacts due to the patient's free breathing during scan are often observable in 4D-CT, which could undermine the procedure of correspondence detection in the registration. Another challenge of motion estimation in 4D-CT is how to keep the lung motion consistent over time. However, the current approaches fail to meet this requirement since they usually register each phase image to a pre-defined phase image independently, without considering the temporal coherence in 4D-CT. To overcome these limitations, we present a unified approach to estimate the respiratory lung motion with two iterative steps. First, we propose a new spatiotemporal registration algorithm to align all phase images of 4D-CT (in low-resolution) onto a high-resolution group-mean image in the common space. The temporal consistency is persevered by introducing the concept of temporal fibers for delineating the spatiotemporal behavior of lung motion along the respiratory phase. Second, the idea of super resolution is utilized to build the group-mean image with more details, by integrating the highly-redundant image information contained in the multiple respiratory phases. Accordingly, by establishing the correspondence of each phase image w.r.t. the high-resolution group-mean image, the difficulty of detecting correspondences between original phase images with missing structures is greatly alleviated, thus more accurate registration results can be achieved. The performance of our proposed 4D motion estimation method has been extensively evaluated on a public lung dataset. In all experiments, our method achieves more accurate and consistent results in lung motion estimation than all other state-of-the-art approaches.

1 Introduction

The modern 4D-CT technique is very useful in many clinical applications, which provides the solid way for the researchers to investigate the dynamics of organ motions in the patient. For example, 4D-CT has been widely used in lung cancer radiation treatment planning for the design of radiation beams to cover the tumor in different respiratory phases. However, the respiratory motion is the one of the major

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 532–539, 2011.

sources of uncertainty in radiotherapy planning of thoracic tumors and many other image guided procedures [1]. Therefore, there is increasing growth in investigating the methods for accurate estimation of the respiratory motion in 4D-CT [2-4].

Image registration plays an important role in the current motion estimation methods by establishing temporal correspondences, e.g., between the maximum inhale phase and all other phases [3, 4]. However, there are some critical limitations in these methods. The first obvious limitation is the independent registration for different pairs of phase images. In this way, the coherence in 4D-CT is totally ignored in the whole motion estimation procedure, thus making it difficult to maintain the temporal consistency along the respiratory phases. The second limitation is mainly coming from the image quality of 4D-CT. Since the modern CT scanner can only scan a limited region of human body at each couch position, the final 4D-CT has to be assembled by sorting multiple free-breathing CT segments w.r.t. the couch position and tidal volume [3]. However, due to the patient's free breathing during scan, no CT segment can be scanned exactly at the particular tidal volume. Thus, motion artifacts can be observed in the 4D-CT, including motion blur, discontinuity of lung vessel, and irregular shape of lung tumor. All these artifacts in 4D-CT challenge the registration methods to establish reasonable correspondences. The final limitation is that the motion estimation in the current methods is generally completed with a single step, i.e., registering all phase images to the fixed phase image, without providing any chance to rectify the possible mis-estimation of respiratory motions.

To overcome these limitations, we present a novel registration-based framework to estimate the lung respiratory motion in 4D-CT. Our method consists of two iterative steps. In the first step, a spatiotemporal registration algorithm is proposed to simultaneously align all phase images onto a group-mean image in common space. Particularly, rather than equally estimating the correspondence for each point, we propose to hierarchically select a set of key points in the group-mean image as the representation of group-mean shape, and let them drive the whole registration process by robust feature matching with each phase image. By mapping the group-mean shape to the domain of each phase image, every point in the group-mean shape has several warped positions in different phases, which can be regarded as a virtual temporal fiber once assembling them into a time sequence. Thus, the temporal consistency of registration is achieved by requiring the smoothness on all temporal fibers during the registration. In the second step, after registering all phase images, the respiratory motion of lung can be obtained to guide the reconstruction of the high-resolution group-mean image. By repeating these two steps, not only more accurate but also more consistent motion estimation along respiratory phases can be achieved, as confirmed in the experiment on a public lung dataset [2], by comparing with the diffeomorphic Demons [5] and the B-spline based 4D registration method [4].

2 Methods

The goal of our method is to estimate the respiratory motion $\varphi(x,t)$ at arbitrary location x and time t. In 4D-CT, only a limited number of phase images are acquired. Thus, given the 4D-CT image $\boldsymbol{I} = \{I_s | s = 1, \dots, N\}$ acquired with N phases, we can model the respiratory motion $\varphi(x,t)$ as a continuous trajectory function with its

known landmark at the position x of the phase t_s, i.e., $\varphi(x, t_s) = f_s(x)$, which is obtained by simultaneously estimating the deformation fields, $\boldsymbol{F} = \{f_s(x)|x \in \Omega_c, s = 1, \ldots, N\}$, from the common space C to all N phase images.

Our method consists of two iterative steps, as shown in Fig. 1. First, we will calculate the deformation field f_s for each phase image I_s w.r.t. the current-estimated group-mean image G in the common space. Generally, G has higher image resolution than all I_s s, since it integrates all information from the aligned I_s s. Thus, the registration can be performed between the high-resolution group-mean image G and the low-resolution phase images I_s, by requiring the respiratory motion $\varphi(x, t)$ to be temporally smooth across all N phases. Second, after aligning all phase images onto the common space, the procedure of *super resolution* (SR) is performed to construct a high-resolution group-mean image G based on the latest estimated deformation fields $\boldsymbol{F}$ and the corresponding phase images $\boldsymbol{I}$. By repeating these two steps, all phase images can be jointly registered onto the common space, with their temporal motions well delineated by $\varphi(x, t)$. The final super-resolution group-mean image G integrates all image information from all phases, thus it has much clearer anatomical details than any phase image I_s, as will be shown next (Fig. 3). Also, it can help resolve the uncertainty in correspondence detection for the case of directly registering different phase images, since, in our method, each point in the phase image I_s can easily find its correspondence in the group-mean image (which has complete image information).

Our method differs from the conventional group-wise registration method [6] in several ways: 1) The registration between group-mean image and all phase images is simultaneously performed by exploring the spatiotemporal behavior on a set of temporal fibers, instead of independently deploying the pairwise registrations; 2) The correspondence is always determined between high-resolution group-mean image and low-resolution phase image; 3) The construction of group-mean image considers the alignment of local structures, instead of simple intensity average, thus better alleviating the artifacts of misalignment. Next, we will explain the proposed spatiotemporal registration and super resolution methods.

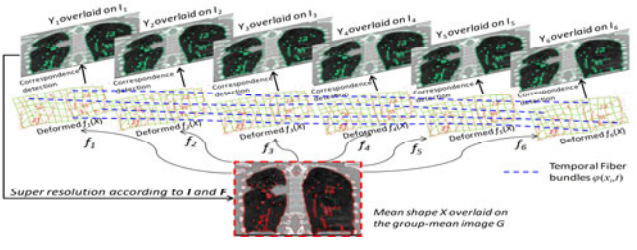


Fig. 1. The overview of our method in respiratory motion estimation by joint spatiotemporal registration and high-resolution image reconstruction

2.1 Hierarchical Spatiotemporal Registration of 4D-CT

Before registration, we will first segment each phase image into bone, soft tissue (muscle and fat), and lung. Then, the vessels inside lung will be enhanced by multi-scale hessian filters [7], in order to allow the registration algorithm to focus on the alignment of lung vessels during the registration. Meanwhile, instead of using only

image intensity, we employ an attribute vector $\vec{a}(x, I_s)$ as the morphological signature for each point x in the image I_s to characterize its local image appearance. The attribute vector consists of image intensity, Canny edge response, and geometric moments of each segmented structure (i.e., bone, soft tissue, lung, and vessel).

In general, simultaneous estimation of deformation fields for all phase images is very complicated and vulnerable of being trapped in local minima during the registration. Since different image points play different roles in registration, we select to focus only on a limited number of key points with distinctive attribute vectors (i.e., obtained by thresholding the Canny edge response in the attribute vector) and let them drive the whole registration. Both key points in the group-mean image, $\boldsymbol{X} = \{x_i | i = 1, \dots, P\}$, and in each phase image I_s, $\boldsymbol{Y}_s = \{y_j^s | j = 1, \dots, Q_s\}$, are extracted, as overlaid in red and green in Fig. 1, respectively, to mutually guide the registration. With use of these key points, we can decouple the complicated registration problem into two simple sub-problems, i.e., (1) robust correspondence detection on the key points and (2) dense deformation interpolation from the key points.

Robust Correspondence Detection by Feature Matching: Inspired by [8], the key points $\boldsymbol{X}$ in the group-mean image G can be considered as the mean shape of the aligned phase images. Similarly, the respective shape in each phase image I_s can be represented by its key point set $\boldsymbol{Y}_s$. Then, each key point y_j^s in the phase image I_s can be regarded as an observation drawn from the Gaussian mixture model (GMM) with its centers as the deformed mean shape $\boldsymbol{X}$ in I_s. In order to find the reliable anatomical correspondences, we also require the attribute vector $\vec{a}(x_i, G)$ of the key point x_i in the group-mean image to be similar with the attribute vector of its corresponding point in I_s. Thus, the discrepancy criterion is given as:

$$\eta(x_i, y_j^s) = \beta \cdot \left\| f_s(x_i) - y_j^s \right\|^2 + (1 - \beta) \cdot \left\| \vec{a}(x_i, G) - \vec{a}(y_j^s, I_s) \right\|^2, \quad (1)$$

where β balances between shape and appearance similarities.

Soft correspondence assignment [8] is further used to improve the robustness of correspondence detection by calculating the matching probability for each candidate. Specifically, we use $\pi_{i,j}^s$ to denote the likelihood of y_j^s being the true correspondence of x_i in the phase image I_s. It is worth noting that the correspondence assignment is hierarchically set for soft correspondence detection in the beginning, and turn to one-to-one correspondence in the final registration in order to increase the specificity of correspondence detection results. Accordingly, the entropy of correspondence assignment $\pi_{i,j}^s$ is set from large to small value with progress of registration. As we will explain later, this dynamic setting is controlled by introducing the temperature r to the entropy degree of $\pi_{i,j}^s$, which is widely used in annealing system [8].

Kernel Regression on Temporal Fibers: After detecting correspondence for each key point x_i, its deformed position $f_s(x_i)$ in each phase image I_s can be regarded as the landmark of $\varphi(x_i, t)$ at time t_s. Here we call this motion function $\varphi(x_i, t)$ on the key point x_i as the *temporal fiber*, with the total number of temporal fibers equal to P (i.e., the number of key points in the group-mean image). In the middle of Fig. 1, we show the deformed group-mean shape in the space of each phase image. By sequentially connecting the deformed position $f_s(x_i)$ along respiratory phases, a set of temporal fibers can be constructed, which are shown as blue dashed curves in Fig. 1.

The advantages of using temporal fibers include: 1) The modeling of temporal motion is much easier on the particular temporal fiber $\varphi(x_i, t)$ than on the entire motion field $\varphi(x, t)$; 2) The spatial correspondence detection and temporal motion regularization are unified along the temporal fibers. Here, we model the motion regularization on each temporal fiber as the kernel regression problem with kernel function :

$$arg\min_{\varphi} L_T\big(\varphi(x,t)\big) = \sum_{s=1}^{N} [\varphi(x,t) - \varphi(x,t_s)]^2 \frac{1}{\sigma}\psi\left(\frac{t-t_s}{\sigma}\right), \tag{2}$$

Energy Function of Spatiotemporal Registration: The energy function in our spatiotemporal registration method is defined as:

$$E = \sum_{s=1}^{N}\left\{\sum_{i=1}^{P}\sum_{j=1}^{Q_s}\left[\pi_{i,j}^s \cdot \eta\big(x_i, y_j^s\big) + r \cdot \pi_{i,j}^s \cdot log\big(\pi_{i,j}^s\big)\right] + \lambda_1 L_s(f_s)\right\} + \lambda_2 \sum_{i=1}^{P} L_t\big(\varphi(x_i,t)\big), \tag{3}$$

where $L_s(f_s)$ is the bending energy for requiring the deformation field f_s to be spatially smooth [8]. λ_1 and λ_2 are the two scalars to balance the strengths of spatial smoothness L_s and temporal consistency L_T. The terms in the square brackets measure the alignment between the group-mean shape $\boldsymbol{X}$ and each shape $\boldsymbol{Y}_s$ in the s-th phase image, where r is used similarly as the temperature in annealing system to dynamically control the correspondence assignment from soft to one-to-one correspondence, as explained next.

Optimization of Energy Function: First, the spatial assignment $\pi_{i,j}^s$ can be calculated by minimizing E in Eq.3 w.r.t. $\pi_{i,j}^s$:

$$\pi_{i,j}^s = exp\left\{-\frac{\eta\big(x_i, y_j^s\big)}{r}\right\} \Big/ \sum_{j=1}^{Q_s} exp\left(-\frac{\eta\big(x_i, y_j^s\big)}{r}\right). \tag{4}$$

It is clear that $\pi_{i,j}^s$ is penalized in the exponential way according to the discrepancy degree $\eta\big(x_i, y_j^s\big)$ defined in Eq. 1. Notice that r is the denominator of the exponential function in Eq. 4. Therefore, when r is very high in the beginning of registration, although the discrepancy between x_i and y_j^s is large, the key point y_j^s still might have the chance to be selected as the correspondence of x_i w.r.t. I_s. In order to increase the registration accuracy, the specificity of correspondence will be encouraged by gradually decreasing the temperature r, until only the key points y_j^s with the smallest discrepancy will be considered as the correspondences of x_i in the end of registration.

Then, the correspondence of each x_i w.r.t. the shape Y_s in the s-th phase image can be determined as the weighted mean location of all y_j^ss, i.e., $f_s(x_i) = \sum_{j=1}^{Q_s} \pi_{i,j}^s y_j^s$, by discarding all unnecessary terms with $f_s(x_i)$ in Eq. 3. Recall that $\varphi(x_i, t_s) = f_s(x_i)$, thus a set of temporal fibers can be constructed to further estimate the continuous motion function $\varphi(x_i, t)$ by performing the kernel regression on a limited number of landmarks $\varphi(x_i, t_s)$. Here, we use $\hat{\varphi}(x_i, t_s)$ to denote the deformed position of x_i in phase image I_s after kernel regression. The last step is to interpolate each dense deformation field f_s. TPS [8] is used to calculate dense deformation field f_s by considering $\{x_i\}$ and $\{\hat{\varphi}(x_i, t_s)\}$ as the source and target point sets, respectively, which has the explicit solution to minimize the bending energy $L_s(f_s)$. In the next section, we will introduce our super-resolution method for updating of the

group-mean image G with guidance of the estimated f_ss. After that, the group-mean shape $\boldsymbol{X}$ can be extracted in the updated group-mean image G again to guide the spatiotemporal registration in next round. We use the 3DCT in the maximum inhale stage as the initialization since its image quality is better than all other phases.

2.2 Construction of High Resolution Group-Mean Image

Given the deformation field f_s of each phase image I_s, all image information of 4D-CT in different phases can be brought into the common space. Thus, the highly-redundant information among all registered phase images can be utilized to reconstruct the high-resolution group-mean image by the technique of super resolution [9]. It is worth noting that we consider both spatiotemporal registration and group-mean construction in the same physical space, although the image resolutions of group-mean image and each phase image are different. Thus, given the point v from low-resolution phase image I_s, we are able to locate its corresponding position in high-resolution space as $f_s^{-1}(y)$, where f_s^{-1} is the inverse deformation field of f_s.

Here, we follow the generalized nonlocal-means approach [9] to estimate $G(x)$ by minimizing the differences of local patches between high-resolution image G and each aligned low-resolution phase image I_s. By taking the possible misalignment during registration into account, the calculation of $G(x)$ is the weighted intensity average of not only all corresponding points $f_s(x)$ in different phase images I_s but also their neighboring points in the neighborhood $n_I(f_s(x))$:

$$G(x) = \sum_{s=1}^{N} \sum_{u \in n_I(f_s(x))} w(x, G; u, I_s) \cdot I_s(u) \Big/ \sum_{s=1}^{N} \sum_{u \in n_I(f_s(x))} w(x, G; u, I_s), \tag{5}$$

where $w(x, G; u, I_s) = \sum_{v \in n_w(u)} \left[1 - \left\| \vec{a}\left(f_s^{-1}(v), G\right) - \vec{a}(v, I_s) \right\|^2\right]$, measuring the alignment of local structures by computing the regional similarity on attribute vectors in neighborhood n_w. In our implementation, the neighborhood size of n_w is fixed (i.e., 5 mm) throughout registration. The size of neighborhood n_I will gradually reduce from 3 mm to only consider the center point in the end of registration.

3 Experiments

To demonstrate the performance of our proposed registration algorithms in estimating lung motions, we evaluate its registration accuracy on DIR-lab data [2], by comparison with the pairwise diffeomorphic Demons [5] and B-spline based 4D registration algorithm [4]. For the same dataset, the computation time for Diffeomorphic Demons and our method is 40min and ~2h (Intel Quad 2.3GHz) respectively, comparable with 1.5h by B-spline based 4D registration method.

Evaluation of Motion Estimation Accuracy on DIR-Lab Dataset: There are 10 cases in DIR-lab dataset, each case having a 4D-CT with six phases. The resolution for intra-slice is around 1mm×1mm, and the thickness is 2.5mm. For each case, 300 corresponding landmarks in the maximum inhale (MI) and the maximum exhale (ME) phases are available with manual delineations, and also the correspondences of 75 landmarks are provided for each phase. Thus, we can evaluate the registration accuracy by measuring the Euclidean distance between the reference landmark positions and

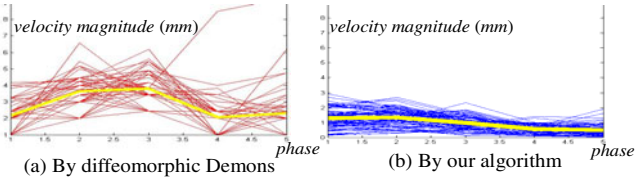
those estimated by the registration method. It is worth noting that our registration is always performed on the entire 4D-CT, regardless in the evaluation of 300 landmarks in MI and ME phases or 75 landmarks in all 6 phases. The image resolution of the group-mean image is increased to 1mm×1mm×1mm.

The registration results by Demons, B-splines based 4D registration, and our algorithm on 300 landmarks between MI and ME phases are shown in the left part of Table 1. Note that we show the results only for the first 5 cases, since the authors in [4] also reported their results only for these 5 cases. It can be observed that our method achieves the lowest mean registration errors. Also, the mean and standard deviation on 75 landmark points over all six phases by the two 4D registration methods, the B-spline based 4D registration algorithm and our algorithm, are shown in the right part of Table 1. Again, our method achieves the lower registration errors.

Fig. 2 (a) and (b) show the velocity magnitudes between consecutive phases for all 75 trajectories, obtained by Demons and our algorithm, respectively. The yellow curves denote the mean velocity magnitude along phases. Since we use the temporal fibers to constrain the continuity along phases, it can be observed that the velocity magnitude is much more continuous by our algorithm.

Table 1. The mean and standard devidation of registration error (mm) on 300 landmark points between maximum inhale and exhale phases, and on 75 landmark points in all six phases

	300 landmark points				*75 landmark points*		
#	*Initial*	*Demons*	*Bspline4D*	*Our method*	*Initial*	*Bspline4D*	*Our Method*
1	3.89(2.78)	2.91(2.34)	1.02 (0.50)	**0.64(0.61)**	2.18(2.54)	0.92 (0.66)	**0.51(0.39)**
2	4.34(3.90)	4.09(3.67)	1.06 (0.56)	**0.56(0.63)**	3.78(3.69)	1.00 (0.62)	**0.47(0.34)**
3	6.94(4.05)	4.21(3.98)	1.19 (0.66)	**0.70(0.68)**	5.05(3.81)	1.14 (0.61)	**0.55(0.32)**
4	9.83(4.85)	4.81(4.26)	1.57 (1.20)	**0.91(0.79)**	6.69(4.72)	1.40 (1.02)	**0.69(0.49)**
5	7.48(5.50)	5.15(4.85)	1.73 (1.49)	**1.10(1.14)**	5.22(4.61)	1.50 (1.31)	**0.82(0.71)**



(a) By diffeomorphic Demons (b) By our algorithm

Fig. 2. The estimated velocity magnitudes from MI to ME phase, by Demons (a) and our algorithm (b), respectively. The yellow curve is the mean velocity magnitude along phases.

Evaluation of High-Resolution Group-Mean Image: After obtaining the high-resolution group-mean image, we can map it back onto the original phase image space. Since the estimated group-mean image has richer information than any phase image in 4D-CT, some missing anatomical structures in the individual phase image could be recovered after mapping our group-mean image with high resolution onto each phase image space. The left panel in Fig. 3 shows the vessel trees at MI, middle phase (MP) between MI and ME, and ME phases of one original 4D-CT. Note that the vessels are extracted by setting threshold (top 10%) on the hessian map and further morphologically processed. Using the same technique, the vessel trees are also extracted from the deformed group-mean images at all phases and shown in the right panel of Fig. 3.

It can be observed that (1) more details on vessels have been recovered by employing the super resolution technique in our method, and (2) the vessel trees are temporally consistent along the respiratory lung motion.

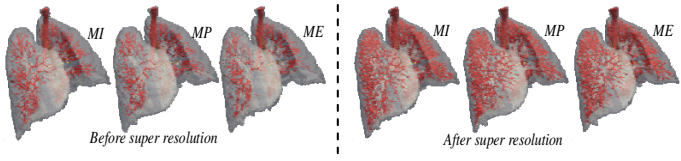


Fig. 3. The vessels tree in all respiratory phases before and after super resolution

4 Conclusion

In this paper, a novel motion estimation method has been presented to measure lung respiratory motion in 4D-CT. Our method is completed by repeating two iterative steps, i.e., (1) simultaneously aligned all phase images onto the common space by spatiotemporal registration and (2) estimating the high-resolution group-mean image with improved anatomical details by super resolution technique. Compared to the other two state-of-the-art registration methods, our algorithm improves not only the estimated motion accuracy, but also the temporal consistency of lung motion.

References

1. Keall, P., Mageras, G., Balter, J., Emery, R., Forster, K., Jiang, S., Kapatoes, J., Low, D., Murphy, M., Murray, B., Ramsey, C., Van Herk, M., Vedam, S., Wong, J., Yorke, E.: The management of respiratory motion in radiation oncology report of AAPM Task Group 76. Med. Phys. 33, 3874–3900 (2006)
2. Castillo, R., Castillo, E., Guerra, R., Johnson, V.E., McPhail, T., Garg, A.K., Guerrero, T.: A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets. Physics in Medicine and Biology 54, 1849–1870 (2009)
3. Ehrhardt, J., Werner, R., Schmidt-Richberg, A., Handels, H.: Statistical Modeling of 4D Respiratory Lung Motion Using Diffeomorphic Image Registration. IEEE Transactions on Medical Imaging 30, 251–265 (2011)
4. Metz, C.T., Klein, S., Schaap, M., van Walsum, T., Niessen, W.J.: Nonrigid registration of dynamic medical imaging data using nD + t B-splines and a groupwise optimization approach. Medical Image Analysis 15, 238–249 (2011)
5. Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Diffeomorphic demons: efficient non-parametric image registration. NeuroImage 45, 61–72 (2009)
6. Joshi, S., Davis, B., Jomier, M., Gerig, G.: Unbiased diffeomorphic atlas construction for computational anatomy. NeuroImage 23, S151–S160 (2004)
7. 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)
8. 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, 160–172 (2004)
9. Protter, M., Elad, M., Takeda, H., Milanfar, P.: Generalizing the Nonlocal-Means to Super-Resolution Reconstruction. IEEE Transactions on Image Processing 18, 36–51 (2008)

Combined Cardiac and Respiratory Motion Compensation for Atrial Fibrillation Ablation Procedures

Alexander Brost[1], Wen Wu[2], Martin Koch[1], Andreas Wimmer[1], Terrence Chen[2], Rui Liao[2], Joachim Hornegger[1], and Norbert Strobel[3]

[1] Pattern Recognition Lab, Friedrich-Alexander-University Erlangen-Nuremberg, Erlangen, Germany
Alexander.Brost@informatik.uni-erlangen.de
[2] Siemens Corporate Research, Princeton, NJ, USA
[3] Siemens AG, Forchheim, Germany

Abstract. Catheter ablation of atrial fibrillation has become an accepted treatment option if a patient no longer responds to or tolerates drug therapy. A main goal is the electrical isolation of the pulmonary veins attached to the left atrium. Catheter ablation may be performed under fluoroscopic image guidance. Due to the rather low soft-tissue contrast of X-ray imaging, the heart is not visible in these images. To overcome this problem, overlay images from pre-operative 3-D volumetric data can be used to add anatomical detail. Unfortunately, this overlay is compromised by respiratory and cardiac motion. In the past, two methods have been proposed to perform motion compensation. The first approach involves tracking of a circumferential mapping catheter placed at an ostium of a pulmonary vein. The second method relies on a motion estimate obtained by localizing an electrode of the coronary sinus (CS) catheter. We propose a new motion compensation scheme which combines these two methods. The effectiveness of the proposed method is verified using 19 real clinical data sets. The motion in the fluoroscopic images was estimated with an overall average error of 0.55 mm by tracking the circumferential mapping catheter. By applying an algorithm involving both the CS catheter and the circumferential mapping catheter, we were able to detect motion of the mapping catheter from one pulmonary vein to another with a false positive rate of 5.8 %.

1 Introduction

Atrial fibrillation (AFib) is widely recognized as a leading cause of stroke [1]. An increasingly popular treatment option for AFib is catheter ablation during which the pulmonary veins attached to the left atrium (LA) are electrically isolated [2]. These procedures are performed in electrophysiology (EP) labs often equipped with modern C-arm X-ray systems providing 3-D imaging of the heart [3]. The use of fluoroscopic overlay images, perspectively rendered from 3-D data sets, makes it possible to augment live 2-D X-ray projections with soft-tissue information [4]. Unfortunately, catheter navigation using augmented fluoroscopy may be

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 540–547, 2011.

(a) (b) (c)

Fig. 1. (a) Cropped image for catheter segmentation. The position of the region is determined by the previous tracking result. (b) Segmented catheter using our boosted classifier approach. (c) The L2-distance transform of the segmentation after skeletonization is used as cost function for our combined registration-tracking approach.

compromised by cardiac and respiratory motion. The first approach to overcome this problem, by providing a motion compensated fluoroscopic overlay image, has been proposed in [5]. It involved tracking of a commonly used circumferential mapping (CFM) catheter firmly positioned at the ostium of the pulmonary vein in simultaneous biplane images. Drawbacks of this method are simultaneous biplane imaging and a need to detect when the CFM catheter has been moved from one PV to another. Another method for respiratory motion compensation has been introduced that tracks a catheter placed in the coronary sinus (CS) vein [6]. Its drawback is that the CS catheter is far away from the site of ablation. As a consequence,it is not obvious that the motion derived from the CS catheter can be directly applied to compensation of both cardiac and respiratory motion. For example, relying on the CS catheter for motion compensation, we may encounter two problems. First, the correlation between the observed CS catheter motion and the actual motion required for dynamic overlay may be low. Second, the catheter in the coronary sinus may not always be visible due to collimation to reduce X-ray dose. Motion compensation using the CS catheter in [6] focused on 2-D motion compensation. Since simultaneous biplane fluoroscopy is hardly used in clinical practice, we also restrict ourselves to 2-D motion compensation as in [7,6]. We present a novel method that combines the best ideas of the previous methods. The CS catheter is used as a point of reference to detect when the circumferential mapping catheter has been moved from one PV to another, while the motion estimate for adapting the fluoroscopic overlay is derived by localizing the CFM catheter. To obtain a good anchor point along the CS catheter, we decided to introduce a virtual electrode (VE). It is placed on the CS catheter more proximal than any other electrode. Below, we briefly explain our new motion compensation method first. Then we turn to our evaluation and the results. Afterwards, we discuss our results and draw conclusions from this work.

2 Motion Compensation

Our motion compensation approach involves tracking of the CFM catheter as well as tracking of a VE placed on the CS catheter. The absolute distance between the center of the CFM catheter and the VE is used to identify whether the CFM has been moved from one PV to another.

2.1 Circumferential Mapping Catheter Tracking

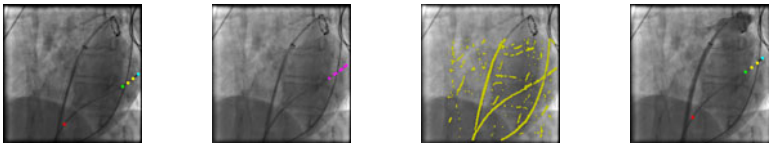
The circumferential mapping catheter is initially extracted in the 2-D X-ray image by manual interaction to get an accurate catheter model. The manual interaction is used only for the first frame of a sequence. The catheter is segmented on a region of interest, see Fig. 1(a). The position of this region is determined by the center of the tracked model from the previous frame. Haar-like features and a cascade of boosted classifiers [8] are used for segmentation. We use classification and regression trees as weak classifiers [9] which are composed to a strong classifier by the AdaBoost algorithm [10]. Our tracking approach follows the method proposed in [7]. A L2-distance map $\boldsymbol{I}_{\mathrm{DT}}$ [11] is calculated from the skeleton of the segmentation, see Fig. 1(c). One pixel $\boldsymbol{p}$ is accessed by $\boldsymbol{I}_{\mathrm{DT}}(\boldsymbol{p})$. Using a 2-D/2-D registration, the result is a 2-D pixel offset $\boldsymbol{t} \in \mathbb{R}^2$. The average distance between model and segmentation derived from the image is considered as the cost value. The optimal translation $\hat{\boldsymbol{t}}$ is found by minimizing

$$\hat{\boldsymbol{t}} = \arg \min_{\boldsymbol{t}} \sum_{i} \boldsymbol{I}_{\mathrm{DT}}(\boldsymbol{q}_i + \boldsymbol{t}) \tag{1}$$

with $\boldsymbol{q}_i$ denoting a point of the catheter model. The estimated 2-D translation $\hat{\boldsymbol{t}}$ is then considered as candidate for motion compensation.

2.2 Coronary Sinus Catheter Tracking

The coronary sinus catheter is modeled by a set of electrodes, starting from the tip of the catheter going through each individual electrode including the most proximal electrode (MPE), to the virtual electrode, see Fig. 2 (a). The VE is a reference point set by clicking on an arbitrary position along the catheter sheath that has to be more proximal than the real MPE. Similar to the circumferential mapping catheter, manual interaction is used to generate the initial CS catheter model in the first frame in a fluoroscopic sequence. For the remaining frames, all the electrodes are tracked using the approach proposed in [12]. Localization of the VE is performed in a two-stage process. In the first stage, we robustly track all the real electrodes. In the second stage, the VE is inferred from the MPE along the CS catheter using a robust tracking framework combined with a geodesic constraint. We rely on learning-based detectors to generate hypotheses of possible models. The classifiers use Haar-like features. Each classifier is a Probabilistic Boosting Tree [13]. To track the VE, specially-designed hypotheses generated by the learning-based catheter body detector are fused. Robust hypothesis matching through a Bayesian framework is then used to select the best hypothesis for each frame. Given the MPE robustly localized in the first stage, the hypotheses for tracking VE are generated automatically by a novel scheme in the second stage. The set of hypotheses is generated by parametrically manipulating the VE model based on the MPE location. A seed hypothesis $\bar{\boldsymbol{v}}_t$ for time step $t \in \mathbb{N}$ is generated by translating $\hat{\boldsymbol{v}}_{t-1}$, a set of 2-D image points in homogeneous coordinates, to the MPE. Then we apply a set of affine transformation to generate tracking hypotheses as



(a) (b) (c) (d)

Fig. 2. CS catheter detection and VE tracking. (a) Fluoroscopic image with user inputs (electrodes and VE). (b) Automatically detected electrode positions after non-max suppression, without decision which electrode is the tip and which electrodes belong to the body. (c) Detected catheter body points. (d) Tracked electrodes and VE.

$$\boldsymbol{v}_t = \begin{pmatrix} \boldsymbol{R} & \boldsymbol{t} \\ 0^T & 1 \end{pmatrix} \cdot \bar{\boldsymbol{v}}_t \tag{2}$$

with $\boldsymbol{R} \in \mathbb{R}^{2\times 2}$ and $\boldsymbol{t} \in \mathbb{R}^2$ being the parameter of an affine transformation. The MPE is used as transformation center. This strategy is efficient in generating a near-complete pool of tracking hypotheses. The overall goal for evaluating a VE tracking hypothesis is to maximize the posterior probability

$$\hat{\boldsymbol{v}}_t = \arg\max_{\boldsymbol{v}_t} P(\boldsymbol{v}_t | \boldsymbol{I}_{0,\ldots,t}) \tag{3}$$

with the image observation $\boldsymbol{I}_{0,\ldots,t}$ from 0 to t-th frame. The tracking is summarized in Fig. 2.

2.3 Displacement Detection

Motion compensation was performed by tracking the circumferential mapping (CFM) catheter. We settled on the CFM approach, because our data did not reveal a sufficiently strong correlation between the motion at the CS catheter and the PV ostium. To explain our findings, let us first recall that the CS catheter, placed in the coronary sinus vein, lies between the left atrium and the left ventricle. As a result, its motion may be highly influenced by the motion of the left ventricle in addition to breathing. The motion of the circumferential mapping catheter, on the other hand, is more constrained because the left atrium is connected to the lungs via the pulmonary veins. Although we decided against using the CS for motion estimation directly, we found it very useful as an anchor, i.e., to detect if the CFM catheter was moved from one PV ostium to the next. To this end, we assumed that the absolute distance between CS catheter and CFM catheter remains sufficiently stable to classify whether the CFM catheter has been moved away from a PV ostium. To achieve a reliable and robust motion compensation, we track both catheters at the same time and compare the absolute 2-D distance between the virtual electrode and the loop's center of the circumferential mapping catheter between two consecutive frames. If the distance

Table 1. Displacement detection using the absolute difference between the CFM catheter and the VE on the CS catheter. False positive (FP) is the percentage of wrongly detected motion and false negative (FN) of undetected motion.

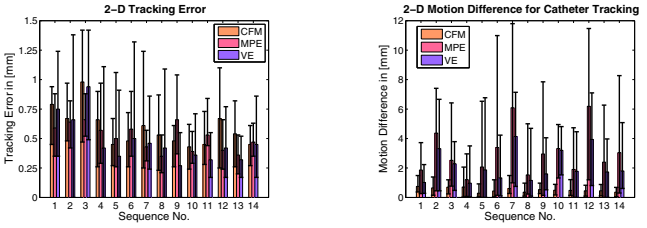
Displacement Detection								
		2 %	5 %	6 %	7 %	10 %	15 %	20 %
VE	*FP*	22.7 %	5.8 %	4.2 %	2.7 %	0.5 %	0.0 %	0.0 %
	FN	0.0 %	0.0 %	14.3 %	57.1 %	42.9 %	57.1 %	85.7 %

changes by more than a certain percentage, we assume that the CFM catheter has been moved from one pulmonary vein to another. In this case, no motion compensation is applied to the fluoroscopic images until another stable CFM position has been reached. Catheter tracking is still performed. As soon as the absolute distance becomes stable again, i.e., the distance change is ≤ 5 %, the motion of the tracked CFM catheter is again applied to the fluoroscopic overlay.

3 Evaluation and Results

Our methods were evaluated on 14 clinical data sets from two different hospitals and from 10 different patients using leave-one-out validation[1]. During three of these sequences, a 10-electrode CS catheter was used. In the remaining data sets, 4-electrode catheters were chosen. The images were either 512×512 pixels or 1024×1024 pixels. The pixel size varied between 0.173 mm and 0.345 mm. Image acquisition was performed without using ECG-triggered fluoroscopy. Hence, both respiratory and cardiac motion were present. At first, we evaluated the accuracy of the tracking methods. The localization error of the detected circumferential mapping catheter was calculated by considering the 2-D Euclidean distance to a gold-standard segmentation. To this end, the L2-distance transform of the gold-standard segmentation was calculated. The 2-D Euclidean distance was then obtained as the average over all values of the distance transform when evaluating it at the 2-D positions of the detected CFM. The gold-standard segmentation was manually generated for each frame in every sequence. The segmentation was supervised by an electrophysiologist. The localization accuracy of the coronary sinus catheter and the virtual electrode was calculated as the 2-D Euclidean distance to a gold-standard segmentation of the whole CS catheter. As before, the L2-distance transform was used. The results are given in Fig. 3(a). The CFM localization yielded an average 2-D error of 0.55 mm, which includes the inherent model error. The detection of the MPE on the CS catheter yielded an average 2-D error of 0.52 mm. The VE detection yielded an average 2-D error of 0.49 mm. Further, we compared the motion calculated from the catheter detection methods to the motion observed at the PV ostia. This motion was obtained by using a gold-standard segmentation of the circumferential mapping catheter. The center of the 2-D catheter model was used to calculate the underlying motion of the

[1] Data is available from the authors on request for non-commercial research purposes.



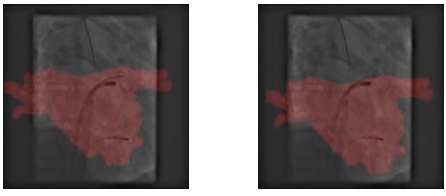
(a) (b)

Fig. 3. (a) Accuracy for the catheter tracking methods. Tracking of the circumferential mapping (CFM) catheter, the most proximal electrode (MPE) on the CS catheter, and the virtual electrode (VE) on the CS catheter. (b) Difference between the observed motion by the circumferential mapping catheter and the catheter tracking methods.

PV between successive frames. The comparison is given in Fig. 3(b). The motion obtained by CFM catheter detection differs on average by about 0.48 mm from the real motion, whereas the motion from the proximal CS electrode had a mean error of about 2.61 mm. Using the virtual electrode, we could reduce the mean error from 2.61 mm to 1.68 mm. The maximum difference between the true and the estimated motion using the CFM catheter was 2.06 mm. The MPE was off by up to 11.80 mm and the VE by up to 7.14 mm, see Fig. 3(b). The 14 fluoroscopic sequences used for evaluating the tracking performance had the CFM catheter firmly placed at a single pulmonary vein, i.e., the CFM catheter was not moved from one PV to the next. To evaluate our displacement detection method, five further sequences were added to our data set. To detect CFM catheter displacement, we introduced a displacement threshold. The displacement threshold is a percentage of the distance between VE and the center of the loop representing the CFM catheter. Results for different displacement thresholds are given in Tab. 1. The best result for displacement detection was found for an allowed change of the distance between 5 % and 10 %. A change in the absolute distance of 5 % turned out to be the best threshold for detecting catheter repositioning in our experiments. In this case, the false positive rate was 5.8 %. We decided on the VE for displacement detection, because it turned out to be a much more stable reference than the MPE. This can be seen, e.g., by taking a look at their mean errors and maximum differences, see Fig. 3(b).

4 Discussion and Conclusions

The results indicate that our catheter localization and tracking algorithms are accurate enough to meet clinical needs, cf. Fig. 3(a). In our experiments, involving non-ECG-triggered X-ray data acquired under free breathing conditions, only tracking of the CFM was accurate enough to be directly applicable to motion compensation without any need for a more sophisticated motion model, see Fig. 3(b). The tracking of the circumferential mapping catheter yielded an



(a) (b)

Fig. 4. (a) Motion compensation using the circumferential mapping catheter. (c) Motion compensation using the most proximal electrode on the coronary sinus catheter.

overall average error of 0.55 mm. Since this error also contains some model error of the underlying B-spline catheter model, which is not adapted over time, the actual tracking performance of the distance-transform-based method is even better. The motion difference between the real motion at the PV ostia and the estimated motion, yielded a maximum error of 2.06 mm. The same error for the MPE was 11.80 mm and 7.14 mm for the VE, respectively, see Fig. 3(b). From these numbers, we conclude that the circumferential mapping catheter is the best surrogate for the motion of the left atrium. A visual comparison between motion compensation using the CFM and the VE is given in Fig. 4. At first sight, our observations seem to contradict the results reported in [6]. Maybe the varying results are due to differences in how the procedures were performed. For example, some centers apply general anesthesia while only mild sedation was used in our cases. Some clinical sites also provide a setup where ECG signals can be recorded on the fluoroscopy system. The ECG could be exploited to select proper fluoroscopic frames. As our cases came from multiple sites using different ECG recording equipment, we decided to not take advantage of any ECG signals to keep things consistent. The choice for one method or the other may come down to how well you control the procedure. For example, if there is general anesthesia, stable sinus rhythm, and available ECG information, the approach presented in [6] may be the method of choice. However, in the general case it may not be straightforward to apply it as successfully. Although we found it difficult to rely on the CS catheter for motion compensation, we observed that it could be used to detect displacement of the CFM catheter. If the distance between the circumferential mapping catheter and the virtual electrode changes by a certain amount, we assume that the mapping catheter has been moved from one PV to the other. From our experiments, using the absolute distance between the CFM and the VE yielded the best results to detect that the CFM moved away from a particular PV. A change in the absolute distance of 5 % was the best threshold in our experiments yielding a false positive rate of 5.8 %. Compared to a misdetection which may lead to incorrect fluoroscopic overlays, a false detection is preferred. At most, there are a few frames without motion correction. As the catheter displacement approach has only been evaluated on five sequences so far, further validation on a more comprehensive data set is currently underway.

References

1. Gage, B., Waterman, A., Shannon, W., Boechler, M., Rich, M., Radford, M.: Validation of Clinical Classification Schemes for Predicting Stroke. Journal of the American Medical Association 285(22), 2864–2870 (2001)
2. Calkins, H., Brugada, J., Packer, D., Cappato, R., Chen, S., Crijns, H., Damiano, R., Davies, D., Haines, D., Haïssaguerre, M., Iesaka, Y., Jackman, W., Jais, P., Kottkamp, H., Kuck, K., Lindsay, B., Marchlinski, F., McCarthy, P., Mont, J., Moradi, F., Nademanee, K., Natale, A., Pappone, C., Prystowsky, E., Raviele, A., Ruskin, J., Shemin, R.: HRS/EHRA/ECAS Expert Consensus Statement on Catheter and Surgical Ablation of Atrial Fibrillation: Recommendations for Personnel, Policy, Procedures and Follow-Up. Heart Rhythm 4(6), 1–46 (2007)
3. Prümmer, M., Hornegger, J.: Cardiac C-arm CT: a unified framework for motion estimation and dynamic CT. IEEE Transactions on Medical Imaging 28(11), 1836–1849 (2009)
4. Ector, J., De Buck, S., Huybrechts, W., Nuyens, D., Dymarkowski, S., Bogaert, J., Maes, F., Heidbüchel, H.: Biplane three-dimensional augmented fluoroscopy as single navigation tool for ablation of atrial fibrillation: Accuracy and clinical value. Heart Rhythm 5(7), 957–964 (2008)
5. 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. LNCS, vol. 5761, pp. 394–401. Springer, Heidelberg (2009)
6. Ma, Y., King, A.P., Gogin, N., Rinaldi, C.A., Gill, J., Razavi, R., Rhode, K.S.: Real-time respiratory motion correction for cardiac electrophysiology procedures using image-based coronary sinus catheter tracking. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6361, pp. 391–399. Springer, Heidelberg (2010)
7. Brost, A., Liao, R., Hornegger, J., Strobel, N.: Model-based registration for motion compensation during EP ablation procedures. In: Fischer, B., Dawant, B.M., Lorenz, C. (eds.) WBIR 2010. LNCS, vol. 6204, pp. 234–245. Springer, Heidelberg (2010)
8. Brost, A., Wimmer, A., Liao, R., Hornegger, J., Strobel, N.: Catheter tracking: Filter-based vs. Learning-based. In: Goesele, M., Roth, S., Kuijper, A., Schiele, B., Schindler, K. (eds.) Pattern Recognition. LNCS, vol. 6376, pp. 293–302. Springer, Heidelberg (2010)
9. Breiman, L., Friedman, J., Olshen, R., Stone, C.: Classification and Regression Trees. Chapman & Hall, New York (1984)
10. Freund, Y., Schapire, R.: A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences 55(1), 119–139 (1997)
11. Breu, H., Gil, J., Kirkpatrick, D., Werman, M.: Linear time Euclidean distance transform algorithms. IEEE Transactions on Pattern Analysis and Machine Intelligence 17(5), 529–533 (1995)
12. Wu, W., Chen, T., Barbu, A., Wang, P., Strobel, N., Zhou, S., Comaniciu, D.: Learning-based Hypothesis Fusion for Robust Catheter Tracking in 2D X-ray Fluoroscopy. To Appear IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2011 (2011)
13. Tu, Z.: Probabilistic boosting-tree: Learning discriminative models for classification, recognition, and clustering. In: Proceedings of the Tenth IEEE International Conference on Computer Vision, ICCV 2005, vol. 2, pp. 1589–1596. IEEE Computer Society, Los Alamitos (2005)

Estimation of Delivered Dose in Radiotherapy: The Influence of Registration Uncertainty

Petter Risholm[1], James Balter[2], and William M. Wells[1]

[1] Brigham and Women's Hospital, Harvard Medical School
[2] Department of Radiation Oncology, University of Michigan
pettri@bwh.harvard.edu

Abstract. We present a probabilistic framework to estimate the accumulated radiation dose and the corresponding dose uncertainty that is delivered to important anatomical structures, e.g. the primary tumor and healthy surrounding organs, during radiotherapy. The dose uncertainty we report is a direct result of uncertainties in the estimates of the deformation which aligns the daily cone-beam CT images with the planning CT. The accumulated radiation dose is an important measure to monitor during treatment, in particular to see if it significantly deviates from the planned dose which might indicate that either the patient was not properly positioned before treatment or that the anatomy has changed due to the treatment. In the case of the latter, the treatment plan should be adaptively changed to align with the current patient anatomy. We estimate the accumulated dose distribution, and its uncertainty, retrospectively on a dataset acquired during treatment of cancer in the neck and show the dose distributions in the form of dose volume histograms.

1 Introduction

Oral and pharyngeal cancers are a significant cause of morbidity and mortality. Approximately 36 000 Americans will be diagnosed with such cancers in 2011, and, of these, only half will survive more than 5 years. Radiotherapy (RT) for cancer in the neck involves extreme precision in planning, image guidance, and delivery, due to the complex distribution of possible tumor locations and their close proximity to a large variety of healthy organs. Using intensity modulated radiation therapy (IMRT), highly sculpted dose distributions can be planned for achieving the dual goals of treating the primary tumor and nearby tissue suspected of containing disease with high doses, while limiting the dose, and thus toxicity risk, to healthy organs.

RT of cancer in the neck region is commonly delivered in 30-35 fractions over 6-7 weeks where the patient must be positioned appropriately for each fraction in order to receive the treatment dose [7]. On any day, the position, orientation, and configuration may vary due to the random nature of the immobilization and setup processes [3]. In addition, chronic changes in response to the treatment and its side effects will manifest as shape changes in the tumor, which typically diminishes in size as treatment progresses, and other tissue changes in the neck

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 548–555, 2011.

due to e.g. weight loss. The typical method employed for ensuring that the dose is delivered safely and effectively on a given day involves changing the position of the patient under image guidance. Such changes, however, are typically restricted to simple translational adjustments [8], as most linear accelerators are not equipped with rotational adjustment systems.

To understand the deviation of dose delivered to a patient from that planned, and thus to guide decisions to adaptively modify treatment plans, it is necessary to accumulate the dose delivered to an important tissue structure as treatment progresses. By non-rigidly aligning the daily Cone-Beam Computed Tomography (CBCT) image with the Computed Tomography (CT) acquired for planning [4], the dose can be summed up for various tissues as treatment progresses and is commonly presented in the form of Dose Volume Histograms (DVH) [1].

The large variety of non-rigid registration methods available have different possible biases as well as random variations due to their driving algorithms and the signals (or degeneracy of signals) available to drive them to optimal solutions. Most methods are deterministic and disregard any concept of uncertainties which may manifest in inaccurate contour mapping and dose reporting. This may lead to erroneous conclusions when deciding whether certain shape changes and their resulting dosimetric impacts warrant more significant intervention than patient position adjustment. In this work we propose a probabilistic framework for estimating the accumulated radiation dose delivered during IMRT in the neck. The daily CBCT is non-rigidly registered with the planning CT using a probabilistic registration framework [5,6] that quantifies the deformation as well as the deformation uncertainty. We show how the delivered dose estimates are influenced by registration uncertainty.

2 Methods

Based on the planning CT image $\mathbf{c}(\mathbf{x})$ and the corresponding binary images of critical structures (e.g. spinal cord, tumor, etc) $\mathbf{s}_1(\mathbf{x}), \ldots, \mathbf{s}_K(\mathbf{x})$, a radiation dose volume $\mathbf{r}(\mathbf{x})$ is determined, where $\mathbf{x} \in \Omega$ and Ω denotes the treatment domain. The total dose is delivered over T identical dose fractions $\mathbf{r}_1 = \ldots = \mathbf{r}_T$ such that $\mathbf{r} = \mathbf{r}_1 + \ldots + \mathbf{r}_T$. Before each fraction, a new CBCT image $\mathbf{f}_t$ is acquired and used to find a translational (rigid) alignment of the patient with respect to the pre-treatment plan.

However, a rigid alignment cannot account for non-rigid anatomical changes, and thus, in those places where non-rigid shifts occur, the dose will not be delivered accurately to the planned treatment site. By estimating the non-rigid deformation $\mathbf{u}_t$ which aligns $\mathbf{f}_t$ with $\mathbf{c}$, we can estimate the deviation between the planned radiation and the actual radiation $\mathbf{d}_t$ delivered to a structure $\mathbf{s}_i$ at fraction t. We model $p(\mathbf{d}_T^*)$, the accumulated marginal dose distribution after fraction T, according to the hierarchical model in Fig. 1 (and drop the conditional dependence of $\mathbf{d}$ on $\mathbf{r}, \mathbf{c}$ and $\mathbf{f}$ for notational convenience):

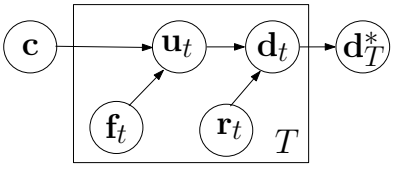


Fig. 1. Hierarchical model of the problem of estimating the actual accumulated radiation dose $\mathbf{d}_T^*$ after T fractions given the planning CT $\mathbf{c}$, the daily CBCT images $\mathbf{f}_t$ and the planned per-fraction dose images $\mathbf{r}_t$. The deformations $\mathbf{u}_t$ and the actual per-fraction doses $\mathbf{d}_t$ are unknowns.

$$p(\mathbf{d}_T^*) = p(\mathbf{d}_1) * \ldots * p(\mathbf{d}_T) \ , \tag{1}$$

$$p(\mathbf{d}_t) = \int_{\mathbf{u}_t} p(\mathbf{d}_t \mid \mathbf{u}_t,\ \mathbf{r}_t)\ p(\mathbf{u}_t \mid \mathbf{c},\ \mathbf{f}_t) \ , \tag{2}$$

where we can estimate $p(\mathbf{d}_T^*)$ by convolution of $p(\mathbf{d}_t)$ for $t = 1, \ldots, T$ because they are conditionally independent given $\mathbf{r}_t$, $\mathbf{c}$ and $\mathbf{f}_t$. In the following sections we show how to compute the posterior distribution over deformations, the conditional dose distribution, the marginal per-fraction dose distribution, and finally the accumulated dose distribution.

2.1 Probabilistic Non-rigid Registration

Here we provide a summary of the probabilistic registration framework in [5,6] where the posterior distribution on deformations is written

$$p(\mathbf{u}_t \mid \mathbf{c},\ \mathbf{f}_t) \propto p(\mathbf{c},\ \mathbf{f}_t \mid \mathbf{u}_t)\ p(\mathbf{u}_t) \ . \tag{3}$$

In the deformation likelihood $p(\mathbf{c},\ \mathbf{f}_t \mid \mathbf{u}_t) = \frac{1}{Z_s} \exp(-E_s/T_s)$ we convert a Sum of Squared Differences (SSD) similarity criterion $E_s(\mathbf{u}_t;\ \mathbf{c},\ \mathbf{f}_t)$ into a probability by way of Boltzmann's equation with temperature T_s. Similarly in the prior $p(\mathbf{u}_t) = \frac{1}{Z_e} \exp(-E_e/T_e)$, Boltzmann's equation with temperature T_e is used to convert a non-linear elastic (St. Venant-Kirchoff) energy $E_e(\mathbf{u}_t)$ into a probability. The computation domain is discretized with a Finite Element (FE) tetrahedral mesh which is used to calculate both E_s and E_e and we prevent folding of elements by setting the elastic energy to infinity when folding is detected.

The method uses Metropolis-Hastings (MH), a Markov Chain Monte Carlo (MCMC) method, to draw deformation samples from the posterior distribution [2]. The MH method constructs a Markov chain that asymptotically has the posterior distribution as its equilibrium distribution. In practice, the method starts with an initial estimate of the deformation $\mathbf{u}^0$, and generates candidate samples from a multi-variate Gaussian proposal distribution centered on the previous sample $\mathbf{u}^* \sim \mathcal{N}(\mathbf{u}^t, \sigma^2 \mathbf{\Sigma})$. The covariance matrix is defined by $\Sigma_{ij} = \exp(-\mathcal{D}(n_i, n_j)/\rho)$ where $\mathcal{D}$ computes the distance between nodes i and j in

the FE-mesh and ρ is a constant. Candidate samples are accepted ($\mathbf{u}^{t+1} = \mathbf{u}^*$), or rejected ($\mathbf{u}^{t+1} = \mathbf{u}^t$), according to the MH criterion [2]. Typically, the first set of samples is discarded to avoid *burn-in* effects, and to reduce intra-sample dependencies the chain is often *thinned* by saving every k-th sample. Eventually, a set of deformations $\hat{\mathbf{u}} = \{\mathbf{u}^i\}_{i=1\ldots N}$ that represent the posterior distribution is generated.

2.2 Marginal Radiation Dose Distributions

We assume dose invariance with respect to anatomical changes and therefore model the conditional per-fraction dose distribution $p(\mathbf{d}_t \mid \mathbf{u}_t,\ \mathbf{r}_t) = \delta(\mathbf{d}_t - \mathbf{r}_t \circ \mathbf{u}_t)$ with a delta distribution. Consequently, we characterize the marginal dose distribution in Eq. (2) by the set of deformed dose samples $\hat{\mathbf{d}}_t = \{\mathbf{r}_t \circ \mathbf{u}^i\}_{i=1\ldots N}$ where $\circ$ denotes composition.

We are interested in monitoring the accumulated dose delivered to a structure $\mathbf{s}_i$ over the duration of the treatment. A simple way to estimate $p(\mathbf{d}_T^*)$, given that we have samples that characterize the per-fraction marginal dose distributions, is by drawing a single random sample from each of the per-fraction dose distributions and adding them together. We draw a set of $M = TNk$ samples $\hat{\mathbf{d}}_T^*$ to characterize $p(\mathbf{d}_T^*)$, where k is a constant that determines the amount of over-sampling we need to get good coverage of the distribution.

A DVH summarizes a 3D dose distribution for a specific structure $\mathbf{s}_i$ in a graphical 2D format and is used for both planning and monitoring radiation treatment [1]. We use it to visualize the most likely radiation dose delivered to a specific structure, either at a single fraction, or the cumulative dose over several fractions. In addition to visualizing the most likely dose distribution, we also visualize the uncertainty (spread) of the dose distribution by plotting the DVHs corresponding to the minimum and maximum limit of the range of the accumulated dose distribution represented by the samples in $\hat{\mathbf{d}}_T^*$.

2.3 Marginal Contour Distributions

It is difficult to assess the final registration results because of the lack of ground truth. We propose to assess the registration through $p(\mathbf{s}_j \circ \mathbf{u}^{-1})$, i.e. the marginal posterior location of the critical structures in the planning image to see if they align with the structure seen in the daily CBCT. It is especially critical to compare the estimated location of these structures with the planned location and the planned dose distribution for those structures where the estimated DVH starts diverging from the planned DVH. A marginal posterior map of the location of the structure in the CBCT can be constructed by deforming the structure with the inverse of each of the sampled deformations $\{\mathbf{s}_j \circ (\mathbf{u}^i)^{-1}\}_{i=1\ldots N}$ and making a histogram of the number of times a voxel is within the structure. We can view this 3D histogram as a probabilistic map which conveys the probability that a voxel belongs to the structure given the posterior distribution over deformations.

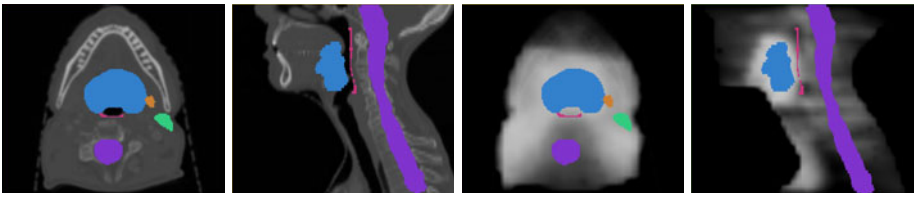


Fig. 2. RT planning data. In the two leftmost images, critical structures (Blue=PTV70, Red=Pharyngeal constrictors, Orange=Left submandibular gland, Green=Left parotid gland, Purple=Spinal cord) used in planning of the radiation dose are overlaid on axial and sagittal slices of the planning CT. In the two rightmost images, corresponding structures are overlaid on axial and sagittal slices of the planned dose $\mathbf{r}$. Notice the steep dose gradients between the spinal cord, pharyngeal constrictor and PTV70.

3 Results

Dataset. We applied our method retrospectively to a Head & Neck data set which consisted of a planning CT (512x512x135) with resolution $(0.95, 0.95, 3.00)$ mm. Important critical structures were delineated on this CT image (see Fig. 2) and a dose volume (61x47x45) with isotropic 5mm resolution was estimated based on the contoured data to maximize the tumor radiation and minimize the radiation to healthy tissue. In the rest of this paper, we will monitor DVHs of normal structures *pharyngeal constrictor*, *brain stem*, *left submandibular gland* and *left parotid gland* to confirm that the actual dose delivered does not exceed the planned dose, and the *planning target volume* (PTV70) to monitor that the delivered dose does not fall below the planned dose.

The patient came in for 35 fractions over 7 weeks, and at each fraction a CBCT (512x512x61) with a resolution of $(0.5, 0.5, 2.5)$mm was acquired to find a translational alignment of the patient with the treatment scan. We used the CBCT from end of week 1, 4 and 7 and assumed that 1/3 of the total dose was applied at each of these time points. We registered the CBCTs to the planning CT to monitor the cumulative DVH for each structure and how they are affected by registration uncertainty and tissue movements due to changing anatomy.

Registration. First we estimated a translation which best aligned the CBCT with the planning CT and assumed this reproduced the translational alignment used during treatment. We generated a FE-mesh, with 1402 nodes and 6398 tetrahedral elements, which covered the treatment region of the neck and used homogeneous elastic parameters with Young's modulus and Poisson's ratio set to $E = 10^8$ and $\nu = 0.45$ respectively. Next we applied our probabilistic registration ($\sigma^2 = 0.008,\ \rho = 50$mm, $T_e = 10^8,\ T_s = 1.0$), sampled 400 000 samples, used a thinning-factor of 10, and removed the first half of the data to avoid burn-in effects. Effectively, we had 20 000 independent samples from the posterior

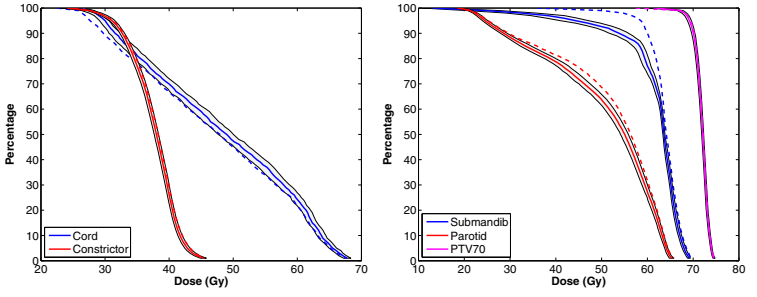


Fig. 3. Cumulative DVH after week 7 for the spinal cord and pharyngeal constrictor on the left and submandibular gland, parotid gland and PTV70 on the right. The gray area is the uncertainty of the DVH and the dashed line is the planned dose. Notice the large deviation from the plan for the submandibular gland and that the planned DVH for the parotid gland falls outside of the estimated DVH. The large deviation from the plan indicates anatomical changes and that the treatment can be replanned. Fig. 4 plots the DVHs for the parotid and submandibular gland from week 1 to 7.

distribution for each dataset. The temperatures were set empirically to balance the contribution of the likelihood and the prior, while σ^2 was set to achieve an MH acceptance rate of approximately 25%.

Dose Estimation. From the per-fraction dose distributions we estimated the cumulative dose distribution per voxel with $k = 100$. In Fig. 3 we show the cumulative DVHs after week 7 and the corresponding uncertainty (dispersion) of the estimates for the 5 different structures we are studying. It is apparent from Fig. 3 that there are deviations in the estimated DVH compared to the planned DVH for the submandibular and parotid glands. In Fig. 4 we study the accumulated dose for these structures closer and plot the cumulative DVH for weeks 1, 4, and 7. After week 1, the actual radiation delivered to the submandibular gland apparently diverges from the planned radiation and is an indication that the tissue in or around the left submandibular gland has moved/changed during treatment and that large dose gradients close to the structure have moved in or out of the structure.

In Fig. 5 we visualize the marginal position of the submandibular and parotid glands overlaid on the week 7 CBCT based on the estimated posterior distribution over deformations to 1) visually validate the registration and 2) to assess if the tissue has moved/changed such that the structure contoured in the treatment CT no longer overlaps with the marginal location of the contour. For the submandibular gland, it is evident that the tissue has moved such that large dose gradients that were previously outside the structure are now inside, which means that the structure is receiving a lower amount than the planned dosage. This might be an indication to re-plan the treatment and possibly taking further advantage of the possible dose sparing to this organ.

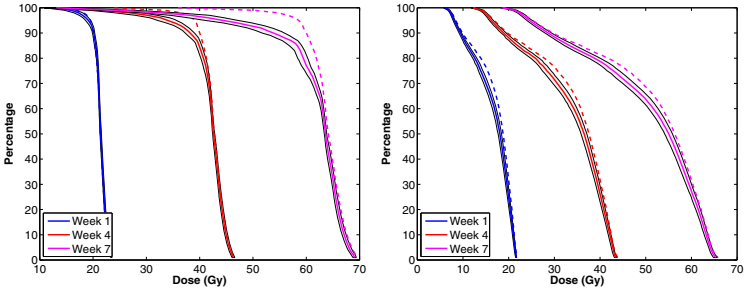
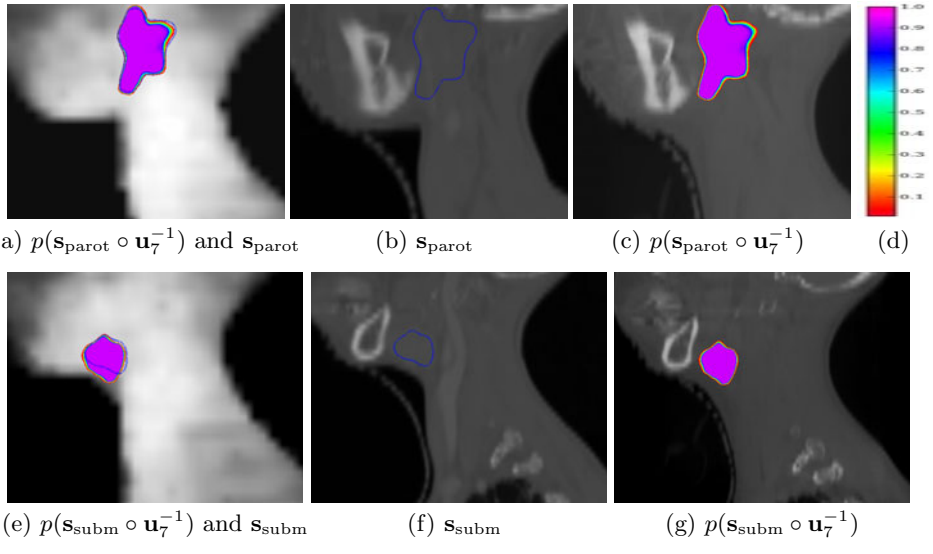


Fig. 4. Cumulative DVHs for the submandibular (left) and the parotid (right) glands. The dashed line is the planned dose, while the gray area is the dose uncertainty. Because we see a large uncertainty in the cumulative DVH for the parotid and submandibular glands in Fig. 3, we plot here the cumulative DVH after weeks 1, 4 and 7 to study the temporal evolution of the delivered dose distributions. At week 4, the estimated DVH for the submandibular gland starts to diverge from the planned DVH. To confirm that these deviations are due to changing and moving anatomy, we plot the marginal distribution of the glands in relation to the planned position in Fig. 5.



(a) $p(\mathbf{s}_{\text{parot}} \circ \mathbf{u}_7^{-1})$ and $\mathbf{s}_{\text{parot}}$ (b) $\mathbf{s}_{\text{parot}}$ (c) $p(\mathbf{s}_{\text{parot}} \circ \mathbf{u}_7^{-1})$ (d)

(e) $p(\mathbf{s}_{\text{subm}} \circ \mathbf{u}_7^{-1})$ and $\mathbf{s}_{\text{subm}}$ (f) $\mathbf{s}_{\text{subm}}$ (g) $p(\mathbf{s}_{\text{subm}} \circ \mathbf{u}_7^{-1})$

Fig. 5. Marginal maps and planned contours of the parotid and submandibular glands. The probability of a voxel being inside the marginal contour is colored according to (d). (a, e): Marginal maps and planned location (in blue) overlaid on $\mathbf{r}$. Notice the deviation of the real location of the submandibular gland (according to the marginal map) from the planned location and how this translates into a deviation in the DVHs (see Fig. 4) because of the large dose gradients in this area. (b, f): Planned contours overlaid on $\mathbf{c}$. (c, g): Marginal maps overlaid on $\mathbf{f}_7$ (week 7 CBCT). Notice the uncertainty in the location of the upper right part of the parotid gland. However, this locational uncertainty does not translate into a large uncertainty in the DVH (see Fig. 3) because there are no large dose gradients in this area.

4 Discussion

This paper presented a probabilistic framework for estimating the cumulative radiation dose delivered to tissue during IMRT and the corresponding dose uncertainty contributed by non-rigid registration. Our preliminary results show that the dose uncertainty varies with the structure we are studying and whether it is close to large dose gradients. The resolution of the dose volume was low which resulted in relatively small dose uncertainties. With finer resolution dose volumes and by accumulating dose from more than three fractions, we anticipate that the dose uncertainties will play a role in deciding whether to re-plan the treatment. The uncertainty estimates also depend on the Boltzmann temperatures. In future work, we will use a likelihood approach to estimate physical meaningful temperatures, and we will also differentiate between tissue types by modeling them with different elasticity. Our current similarity metric (SSD) is not a robust measure, but was used for its computational efficiency. MCMC sampling is generally slow, and we achieved a sampling frequency of 10 samples per second on a 8-core PC for a total computation time of about 11h for one fraction. However, these computations can be done offline between fractions and computational speed is therefore not critical.

Acknowledgments. The work was supported by NIH grants P01CA59827, P41RR13218 and P41RR019703.

References

1. Drzymala, R., Mohan, R., Brewster, L., Chu, J., Goitein, M., Harms, W., Urie, M.: Dose-volume histograms. Int. J. Radiat. Oncol. Biol. Phys. 21(1), 71–78 (1991)
2. Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B.: Bayesian Data Analysis, 2nd edn. Chapman & Hall/CRC (July 2003)
3. Hong, T.S., Tome, W.A., Chappell, R.J., Chinnaiyan, P., Mehta, M.P., Harari, P.M.: The impact of daily setup variations on head-and-neck intensity-modulated radiation therapy. Int. J. Radiat. Oncol. Biol. Phys. 61(3), 779–788 (2005)
4. Lu, W., Olivera, G.H., Chen, Q., Ruchala, K.J., Haimerl, J., Meeks, S.L., Langen, K.M., Kupelian, P.A.: Deformable registration of the planning image (kvct) and the daily images (mvct) for adaptive radiation therapy. Phys. Med. Biol. 51(17), 4357 (2006)
5. Risholm, P., Pieper, S., Samset, E., Wells III, W.M.: Summarizing and visualizing uncertainty in non-rigid registration. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6362, pp. 554–561. Springer, Heidelberg (2010)
6. Risholm, P., Samset, E., Wells III, W.: Bayesian estimation of deformation and elastic parameters in non-rigid registration. In: Fischer, B., Dawant, B.M., Lorenz, C. (eds.) WBIR 2010. LNCS, vol. 6204, pp. 104–115. Springer, Heidelberg (2010)
7. Schwartz, D.L., Dong, L.: Adaptive radiation therapy for head and neck cancer-can an old goal evolve into a new standard? J. Oncol. (2011)
8. Sykes, J.R., Brettle, D.S., Magee, D.R., Thwaites, D.I.: Investigation of uncertainties in image registration of cone beam ct to ct on an image-guided radiotherapy system. Phys. Med. and Biol. 54(24), 7263 (2009)

Deformable Registration of High-Resolution and Cine MR Tongue Images

Jonghye Woo[1,2], Maureen Stone[1], and Jerry L. Prince[2]

[1] Department of Neural and Pain Science, University of Maryland Dental School, Baltimore MD, USA
[2] Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore MD, USA
{jwoo,mstone}@umaryland.edu, prince@jhu.edu

Abstract. This work investigates a novel 3D multimodal deformable registration method to align high-resolution magnetic resonance imaging (MRI) with cine MRI of the tongue for better visual and motion analysis. Both modalities have different strengths to characterize and analyze the tongue structure or motion. Visual and motion analysis of combined anatomical and temporal information can synergistically improve the utility of each modality. An automated multimodal registration method is presented utilizing structural information computed from the 3D Harris operator to encode spatial and geometric cues into the computation of mutual information. The robustness and accuracy of the proposed method have been demonstrated using experiments on clinical datasets and yielded better performance compared to the conventional method and an average error comparable to the inter-observer variability.

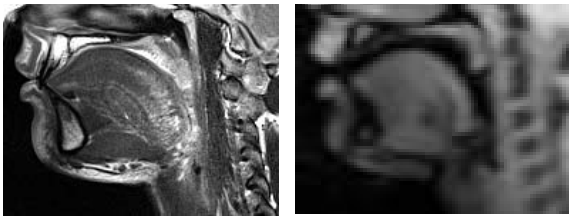
1 Introduction

Assessment of tongue motion can help the early diagnosis of disease, the evaluation of speech quality before and after surgery, and the functional analysis of the tongue [1]. Tongue anatomy is unusual; the tongue has three orthogonal fiber directions and extensive fiber inter-digitation, no bones or joints. This architecture makes the motion pattern of the tongue difficult to measure and quantify.

Assessment, diagnosis, and treatment of tongue disorders and understanding the tongue's motor control can be improved through a combinatorial analysis of tongue muscle anatomy and related tissue motion observed in magnetic resonance (MR) images [2,3]. For example, high-resolution magnetic resonance imaging (hMRI) provides muscle anatomy as shown in Figure 1(a) and cine MRI provides tongue surface motion as shown in Figure 1(b). The combination of hMRI and cine MRI offers complementary information in the study of tongue motion. However, each modality has its limits. hMRI is restricted to a static position and cine MRI does not have sufficient spatial resolution to provide high-quality tongue anatomy. To enhance the advantages of both modalities, it is necessary to combine them through registration.

In this work, we develop a fully automated and accurate 3D deformable registration method to align hMRI with cine MRI. Anatomical (hMRI) and temporal

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 556–563, 2011.



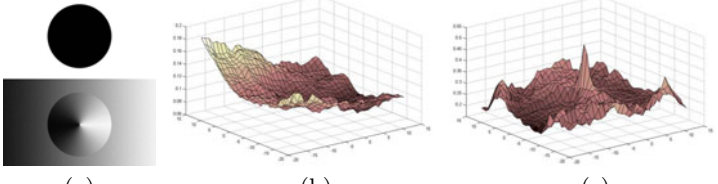
(a) (b)

Fig. 1. Example of the high-resolution MRI (hMRI) that was acquired at rest (a) and the first time frame of cine MRI that was acquired during speech task of "a geese" (b)

(cine MRI) data can be registered to provide correspondences between muscle anatomy identified in hMRI and tongue surface motions in cine, thus mapping changes in muscle pattern with surface motion. To our knowledge, this is the first study to perform registration between these two modalities.

Although mutual information (MI) is considered as a gold standard similarity measure for multimodal image registration, there are two problems in the conventional MI-based registration method. First, MI cannot handle the local intensity variations, which affects the estimation of joint histogram in MI computation [4,5]. Second, the statistics that are computed from overlap regions only considers corresponding intensity information and thus cannot encode spatial information [6]. Figure 2 illustrates a simple yet demonstrative example of this problem. Aligned synthetic circles as in [7] are used to show the limitation of the conventional MI. We compute the cost values with respect to different translations where no local maximum is found in conventional MI as shown in Figure 2(b) whereas the proposed method coincides with a local maximum corresponding to correct alignment as illustrated in Figure 2(c).

These problems were addressed partly by incorporating spatial information into calculation of the MI. Pluim *et al.* [6] combined spatial information by multiplying the MI with an external local gradient. Rueckert *et al.* [8] proposed higher-order mutual information. Russakoff *et al.* [9] proposed regional mutual information to take neighboring information of corresponding pixels into account. Yi *et al.* [7] proposed to include spatial variability via a weighted



(a) (b) (c)

Fig. 2. Comparison between conventional mutual information (b) and the proposed method (c). With aligned synthetic circles (a), we plot the cost values with respect to the different translations along x and y axes.

combination of normalized mutual information and local matching statistics. Zhuang *et al.* [5] proposed to unify spatial information into the computation of the joint histogram. Loeckx *et al.* [10] investigated the conditional MI by incorporating both intensity dimensions and spatial dimension to express the location of the joint intensity pair.

To mitigate the limitations of conventional MI, we propose a novel mechanism to incorporate both spatial and geometric information into the calculation of MI using the Harris operator.

2 Method

Our method is based on an iterative framework of computing MI incorporating spatial information and geometric cues. The underlying idea is to split the image into a set of non-overlapping regions using a 3D Harris operator and to perform registration on spatially meaningful regions. Additionally, we exploit structural information describing gradient of the local neighborhood of each pixel to define structural saliency to compute MI.

2.1 Maximization of Mutual Information

We denote the images $I_1 : \Omega_1 \subset \mathbb{R}^n \rightarrow \mathbb{R}$ and $I_2 : \Omega_2 \subset \mathbb{R}^n \rightarrow \mathbb{R}$, defined on the open and bounded domains Ω_1 and Ω, as the template and target images, respectively. Given two images, a deformation field is defined by the mapping $u : \Omega_2 \mapsto \Omega_1$. The goal of registration is to find a deformation field at each pixel location $\mathbf{x}$ such that the deformed template $I_1(u(\mathbf{x}))$ is as close as possible to $I_2(\mathbf{x})$ satisfying the given criterion. Since I_1 and I_2 are considered to be different modalities, we focus on the MI criterion for registration [4]. The main idea is to find the deformation field $\hat{u}$ by maximizing the statistical dependency between the intensity distributions of the two images, i.e.,

$$\hat{u} = \arg\max_{u}(\mathcal{M}(I_1(u(\mathbf{x})), I_2(\mathbf{x}))), \tag{1}$$

where $\mathcal{M}$ denotes the mutual information of the two distributions. $\mathcal{M}$ can be computed using joint entropy $\mathcal{H}$ as

$$\begin{aligned} \mathcal{H}(I_1(u(\mathbf{x})), I_2(\mathbf{x})) &= -\iint p(i_1, i_2) \log p(i_1, i_2) di_1 i_2 \\ \mathcal{M}(I_1(u(\mathbf{x})), I_2(\mathbf{x})) &= \mathcal{H}(I_1(u(\mathbf{x}))) + \mathcal{H}(I_2(\mathbf{x})) - \mathcal{H}(I_1(u(\mathbf{x})), I_2(\mathbf{x})) \\ &= \int_{\mathbb{R}^3} p_u(i_1, i_2) \log \frac{p_u(i_1, i_2)}{p_{I_1}(i_1) p_{I_2}(i_2)} di_1 di_2, \end{aligned} \tag{2}$$

where $i_1 = I_1(u(\mathbf{x})), i_2 = I_2(\mathbf{x})$, and $p_{I_1}(i_1)$ and $p_{I_2}(i_2)$ are marginal distributions. $p_u(i_1, i_2)$ denotes the joint distribution of $I_1(u(\mathbf{x}))$ and $I_2(\mathbf{x})$ in the overlap region $V = u^{-1}(\Omega_1) \cap \Omega_2$ which can be computed using the Parzen window given by

$$p_u(i_1, i_2) = \frac{1}{|V|} \int_V \varphi\left(\frac{i_1 - I_1(u(\mathbf{x}))}{\rho}\right) \varphi\left(\frac{i_2 - I_2(\mathbf{x})}{\rho}\right) d\mathbf{x}\,, \tag{3}$$

where φ is a Gaussian kernel and ρ controls the width of window.

2.2 Volume Labeling Using 3D Harris Operator

The Harris corner detector [11] was first introduced to detect corner features that contain high intensity changes in the horizontal and vertical directions. In this work, we extend the 2D Harris detector used for images or video sequences to localize meaningful features in 3D images. The Harris operator is based on the local autocorrelation function of the intensity, which measures the local changes of the intensity with patches shifted in different directions. We first define the autocorrelation function as

$$c(x,y,z) = \sum_{x_i,y_i,z_i} W(x_i,y_i,z_i)\left[I(x_i,y_i,z_i) - I(x_i+\Delta x, y_i+\Delta y, z_i+\Delta z)\right]^2, \tag{4}$$

where $I(\cdot,\cdot,\cdot)$ denotes the image function, (x_i, y_i, z_i) are the points in the Gaussian function $W(\cdot,\cdot,\cdot)$ centered on (x,y,z) and $(\Delta x, \Delta y, \Delta z)$ represents a shift to define the neighborhood area. Using a first-order Taylor expansion, we can write

$$\begin{aligned} c(x,y,z) &= \sum_{x_i,y_i,z_i} \left[W \cdot I(x_i+\Delta x, y_i+\Delta y, z_i+\Delta z) \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} \right]^2 \\ &= [\Delta x\, \Delta y\, \Delta z] \begin{bmatrix} \sum\limits_{x_i,y_i,z_i} W \cdot I_x^2 & \sum\limits_{x_i,y_i,z_i} W \cdot I_x I_y & \sum\limits_{x_i,y_i,z_i} W \cdot I_x I_z \\ \sum\limits_{x_i,y_i,z_i} W \cdot I_x I_y & \sum\limits_{x_i,y_i,z_i} W \cdot I_y^2 & \sum\limits_{x_i,y_i,z_i} W \cdot I_y I_z \\ \sum\limits_{x_i,y_i,z_i} W \cdot I_x I_z & \sum\limits_{x_i,y_i,z_i} W \cdot I_y I_z & \sum\limits_{x_i,y_i,z_i} W \cdot I_z^2 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} \\ &= [\Delta x\, \Delta y\, \Delta z]\, \mathcal{C}(x,y,z) \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix}, \end{aligned} \tag{5}$$

where I_x, I_y, and I_z denote the partial derivatives in the x, y, and z axes, respectively, and the *local structure matrix* $\mathcal{C}(x,y,z)$ captures the intensity structure of the local neighborhood. Let $\lambda_1 \geqslant \lambda_2 \geqslant \lambda_3$ denote the eigenvalues of the matrix $\mathcal{C}(x,y,z)$ and let the 3D Harris operator be defined as

$$H = \det(\mathcal{C}) - k(\mathrm{trace}(\mathcal{C}))^3, \tag{6}$$

where k is an arbitrary constant. Then each pixel can be classified as one of three types using a threshold T and the following definitions

- Type 1: $H \geq T$, Location having significant local variation
- Type 2: $H \leq -T$, Location having moderate local variation
- Type 3: $-T \leq H \leq T$, Location having small local variation

We assume that Type 1 and Type 2 regions have more structural and characteristic information compared to Type 3 (homogeneous) region to calculate local statistics. Thus we consider Type 1 and Type 2 regions to calculate MI. One example result of the voxel labeling is shown in Figure 3(b). The white, gray and black color represent the Type 1, Type 2, and Type 3, respectively.

2.3 Mutual Information Using Local Structure Matrix

MI represents the statistical relationship between the template and target images. As shown in Eq. (2), MI is calculated using the marginal and joint distributions of the two images. To address limitations stated before, we compute a weighted joint distribution in order to encode both spatial and geometric information in the objective function. The local structure matrix $\mathcal{C}(x, y, z)$ derived in Eq. (5) exhibits local intensity structure that implies gradient directions within a local neighborhood of each pixel. We can rewrite the joint distribution weighted by the distance between two matrices defined in corresponding pixels as:

$$p_u^{\mathcal{C}}(i_1, i_2) = \frac{1}{|V|} \int_V \gamma(\mathbf{x}) \cdot \varphi\left(\frac{i_1 - I_1(u(\mathbf{x}))}{\rho}\right) \varphi\left(\frac{i_2 - I_2(\mathrm{x})}{\rho}\right) d\mathbf{x} \tag{7}$$

where $\gamma(\mathbf{x})$ is a weighting function that incorporates the distance between local structure matrices between corresponding pixels given by

$$\gamma(\mathbf{x}) = \exp(-\frac{\Delta(\mathcal{C}_{i_1}(\mathbf{x}), \mathcal{C}_{i_2}(\mathbf{x}))}{m}). \tag{8}$$

Here, $\Delta(\mathcal{C}_{i_1}(\mathbf{x}), \mathcal{C}_{i_2}(\mathbf{x}))$ is a distance between two matrices, m is a normalization constant, and $\mathcal{C}_{i_1}(\mathbf{x})$ and $\mathcal{C}_{i_2}(\mathbf{x})$ are the local structure matrices of the corresponding pixels in $I_1(u(\mathbf{x}))$ and $I_2(\mathbf{x})$, respectively. The local structure matrices do not reside in a vector space and therefore the Euclidean metric does not hold. However, local structure matrices are symmetric and positive semidefinite (like covariance matrices), and therefore belong to a connected Riemannian manifold that is locally Euclidean [12]. Accordingly, we can define the distance between two structure matrices as

$$\Delta(\mathcal{C}_{i_1}(\mathbf{x}), \mathcal{C}_{i_2}(\mathbf{x})) = \sqrt{\sum_{n=1}^{N} \ln^2 \lambda_n(\mathcal{C}_{i_1}(\mathbf{x}), \mathcal{C}_{i_2}(\mathbf{x}))}, \tag{9}$$

where λ_n are the generalized eigenvalues of $\mathcal{C}_{i_1}(\mathbf{x})$ and $\mathcal{C}_{i_2}(\mathbf{x})$ and N is the number of rows and columns in each matrix. This definition of distance satisfies the metric properties including symmetry, positivity, and the triangle inequality.

We can rewrite MI based on the above weighting scheme as follows:

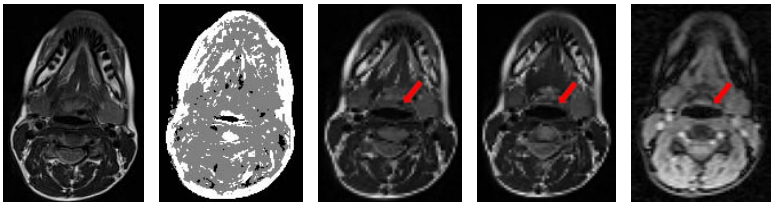
$$\mathcal{M}^{\mathcal{C}}(I_1(u(\mathbf{x})), I_2(\mathbf{x})) = \int_{R^3} p_u^{\mathcal{C}}(i_1, i_2) \log \frac{p_u^{\mathcal{C}}(i_1, i_2)}{p_{I_1}(i_1) p_{I_2}(i_2)} di_1 di_2. \tag{10}$$

Using this (modified) MI, the local structure matrices provide a geometric similarity measure while the image intensities continue to provide an appearance measure, thereby allowing us to find correspondence more reliably and address the limitation of the conventional MI-based registration.

2.4 Registration Model

Data fidelity Term. With the modified MI defined in Eq. (10) and labeled regional information, we can define data fidelity term given by

$$\mathcal{D}(I_1(u(\mathbf{x})), I_2(\mathbf{x})) = \sum_{k=1}^{K} w_k \chi_{D_k}(u(\mathbf{x})) \mathcal{M}^{\mathcal{C}}(I_1(u(\mathbf{x})), I_2(\mathbf{x})), \tag{11}$$



(a) (b) (c) (d) (e)

Fig. 3. One example of the results: (a) template image (hMRI) (b) volume labeling of the template image using the Harris operator, (c) a resulting image using conventional MI-based registration, (d) a resulting image using the proposed method and (e) the target image (cine MRI). The red arrows show that (d) and (e) are better aligned than (c) and (e) in terms of vocal tract edge.

where $w_k \in \mathbb{R}^+$ is the weight assigned to kth region and χ_{D_k} is kth characteristic function defined by

$$\chi_{D_k}(x) = \begin{cases} 1, & x \in D_k \\ 0, & x \notin D_k \end{cases} \tag{12}$$

Transformation Model. We use free-form deformations (FFD) based on uniform cubic B-splines to model the deformable registration as in [13]. Additionally, a multi-resolution scheme is used to represent coarse-to-fine details of both volumes for fast and robust registration. The energy functional is minimized using a Simultaneous Perturbation Stochastic Approximation (SPSA) [14] method.

3 Experiments and Results

3.1 Subjects and Task

Nine normal native American English speakers were subjects in this experiment. The speech task was "a geese". Both types of MRI datasets were recorded in the same session using a head and neck coil. Cine MRI datasets were collected with a 6mm slice thickness and had an in-plane resolution of 1.875mm/pixel resolution. hMRI datasets were 3mm thick with an in-plane resolution of 0.94mm/pixel. The subjects were required to remain still from 1.5 to 3 minutes for each plane.

3.2 Evaluation of the Registration Method

To evaluate the accuracy and robustness of the proposed method, we have performed two registration experiments on nine pairs of 3D axial MRI volumes described above. Both registrations were performed on the same two static volumes: (1) the first time frame of axial cine MRI that was acquired during speech task of "a geese" and (2) the axial hMRI volume that was acquired at rest. The registration methods used affine registration as an initialization, followed by the deformable registration using the proposed and conventional MI-based method using FFD. In our experiments, we set the number of histograms to 50, and used

Table 1. Registration errors and observer variability

TRE (voxel)	Before	Affine	Conventional	Proposed	Observer Variability
Tongue Tip	6.2±3.7	3.8±1.4	3.6±1.8	2.5±1.2	1.8±1.3
Lower Tip	3.9±1.8	2.8±1.1	2.6±1.4	1.8±1.2	2.7±1.7
Posterior pharynx	3.9±5.8	1.9±0.9	1.5±0.7	1.5±0.9	1.4±1.3
Average	4.7±4.0	3.1±2.8	2.7±2.6	2.1±1.2	2.0±1.4

Table 2. Registration errors in different non-uniformity fields

TRE (voxel)	Affine	Conventional method	Proposed method
Small bias field (20%)	3.8±1.6	3.5±2.6	2.3±1.2
Medium bias field (40%)	3.7±1.3	3.6±2.6	2.4±1.2
Large bias field (60%)	3.8±1.5	3.8±2.5	2.7±1.5

the entire volume as the sample size. We used control point spacings of 8 mm in each axis. For the 3D Harris operator, we set k=0.001 and T=50,000,000. The method stops when the movement is less than 0.001 mm or iteration reaches the predefined iteration number 100 in both methods.

The first experiment assessed the accuracy of the registration method using target registration error (TRE) [15]. Two expert observers independently selected three corresponding anatomical landmarks from each volume including tongue tip, lower lip, and posterior pharynx. Table 1 lists the mean and standard deviation of TRE and inter-observer variability using both methods. The TRE results show that the proposed method provides accurate results compared to the conventional MI-based method and is comparable to the observer-variability. Figure 3 shows one result of the first experiment. It is apparent in the figure that the proposed method has better alignment. Of note, selecting anatomical landmarks is of great importance, and a challenging task even for humans, in assessing the accuracy of the registration method. There is no true gold standard other than visual judgment, which is marred by inter-observer variability.

The second experiment further demonstrated the performance of the registration method. Three different levels of intensity non-uniformity (bias) were generated including small (20%), medium (40%) and large (60%) bias fields. In these experiments, we also used TRE to measure the performance of the methods. As shown in Table 2, the results of the proposed method were superior to the conventional method and were also robust against the bias fields.

4 Conclusion

In this work, we propose a novel registration algorithm to align hMRI with cine MRI. We utilize structural information computed from the 3D Harris operator to encode spatial and geometric cues into the computation of MI. Fully automated 3D deformable registration of hMRI with cine MRI of tongue can be performed

accurately with average error of TRE comparable to inter-observer variability. The proposed approach can be applied to the mapping of muscle anatomy in hMRI to tongue surface motions in cine MRI.

References

1. Stone, M., Davis, E., Douglas, A., NessAiver, M., Gullapalli, R., Levine, W., Lundberg, A.: Modeling the motion of the internal tongue from tagged cine-MRI images. The Journal of the Acoustical Society of America 109(6), 2974–2982 (2001)
2. Narayanan, S., Nayak, K., Lee, S., Sethy, A., Byrd, D.: An approach to real-time magnetic resonance imaging for speech production. The Journal of the Acoustical Society of America 115(4), 1771–1776 (2004)
3. Parthasarathy, V., Prince, J., Stone, M., Murano, E., NessAiver, M.: Measuring tongue motion from tagged cine-MRI using harmonic phase (HARP) processing. The Journal of the Acoustical Society of America 121(1)
4. Pluim, J., Maintz, J., Viergever, M.: Mutual-information-based registration of medical images: a survey. IEEE Trans on Medical Imaging 22(8), 986–1004 (2003)
5. Zhuang, X., Hawkes, D.J., Ourselin, S.: Unifying encoding of spatial information in mutual information for nonrigid registration. In: Prince, J.L., Pham, D.L., Myers, K.J. (eds.) IPMI 2009. LNCS, vol. 5636, pp. 491–502. Springer, Heidelberg (2009)
6. Pluim, J.P.W., Maintz, J.B.A., Viergever, M.A.: Image registration by maximization of combined mutual information and gradient information. In: Delp, S.L., DiGoia, A.M., Jaramaz, B. (eds.) MICCAI 2000. LNCS, vol. 1935, pp. 452–461. Springer, Heidelberg (2000)
7. Yi, Z., Soatto, S.: Nonrigid registration combining global and local statistics. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 2200–2207 (2009)
8. Rueckert, D., Clarkson, M., Hill, D., Hawkes, D.: Non-rigid registration using higher-order mutual information. In: Proceedings of SPIE, vol. 3979, p. 438 (2000)
9. Russakoff, D.B., Tomasi, C., Rohlfing, T., Maurer Jr., C.R.: Image similarity using mutual information of regions. In: Pajdla, T., Matas, J(G.) (eds.) ECCV 2004. LNCS, vol. 3023, pp. 596–607. Springer, Heidelberg (2004)
10. Loeckx, D., Slagmolen, P., Maes, F., Vandermeulen, D., Suetens, P.: Nonrigid image registration using conditional mutual information. IEEE Trans on Medical Imaging 29(1), 19–29 (2009)
11. Harris, C., Stephens, M.: A combined corner and edge detector. In: Alvey Vision Conference, Manchester, UK, pp. 147–151 (1988)
12. Donoser, M., Urschler, M., Hirzer, M., Bischof, H.: Saliency driven total variation segmentation. In: CVPR, pp. 817–824 (2010)
13. 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 Trans. Medical Imaging 18(8), 712–721 (1999)
14. Spall, J.: An overview of the simultaneous perturbation method for efficient optimization. Johns Hopkins APL Technical Digest 19(4), 482–492 (1998)
15. Fitzpatrick, J.M., West, J.B., Maurer, C.R.: Predicting error in rigid-body point-based registration. IEEE Trans. Medical Imaging 17, 694–702 (1998)

Motion Compensated Magnetic Resonance Reconstruction Using Inverse-Consistent Deformable Registration: Application to Real-Time Cine Imaging*

Hui Xue[1], Yu Ding[2], Christoph Guetter[1], Marie-Pierre Jolly[1], Jens Guehring[4], Sven Zuehlsdorff[3], and Orlando P. Simonetti[2]

[1] Imaging Analytics and Informatics, Siemens Corporate Research, Princeton, NJ, USA
[2] Davis Heart and Lung Research Institute, The Ohio State University, OH, USA
[3] CMR R&D, Siemens Medical Solutions USA, Inc., Chicago, IL, USA
[4] Imaging & IT Division, Siemens AG, Healthcare Sector, Erlangen, Germany
hui-xue@siemens.com

Abstract. Patient motion is a major limitation for magnetic resonance imaging. Recent theoretical advances incorporate explicit rigid and non-rigid motion compensation into conventional image reconstruction for multi-shot acquisitions and recover motion-free images by solving a general matrix inversion problem. Although the theory has been established, applications are rare due to the challenges of estimating motion field for every pixel of every shot. In this paper we propose a method to overcome this difficulty using the inverse-consistent deformable registration supplying both forward and backward deformations for matrix inversion. We further extend this framework for multi-coil motion compensated image reconstruction using the eigen-mode analysis. Both simulations and in vivo studies demonstrate the effectiveness of our approach.

1 Introduction

Patient motion during magnetic resonance imaging (MRI) causes blurring or ghosting that degrades image quality. The undesired effects of patient motion are introduced during the acquisition of Fourier domain or k-space data. Any inconsistencies in k-space can strongly influence every pixel in the image domain. Even slight motion will cause blurring in linearly acquired k-space or ghosting artifacts in interleaved k-space acquisitions. Significant movement may create artificial structures that may interfere with diagnostic interpretation of the image (Fig. 1).

Motion compensation in MRI is an active research area [1]. Published methods can be classified as either prospective or retrospective. Prospective approaches such as breath-holding, ECG synchronization or respiratory gating require additional clinical set-up, rely on patient cooperation, regular breathing and/or cardiac rhythm, and can lengthen the scan time. The majority of retrospective methods deals with rigid motion, and may be insufficient for applications like cardiac or liver imaging where significant deformation of organs could occur in addition to rigid motion.

Recently proposed methods aim to extend conventional MR reconstruction to incorporate explicit motion compensation for non-rigid deformations. Recent theoretical

* This work was supported in part by the NIH grant RO1 HL102450.

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 564–572, 2011.

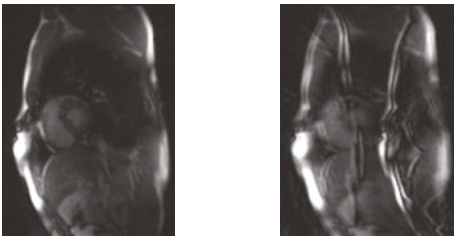


Fig. 1. An illustration of motion induced artifacts. The left shows a cardiac cine frame whose k-space was divided into 2 shots by picking every the other line. Significant artifacts are induced on the right after two shots are shifted by 3 pixels.

advances [2] elucidate that if ***a)*** the MR data acquisition is performed in a multi-shot or multi-segment manner, which is often the case for many cardiac and 3D imaging applications, and ***b)*** the motion between every shot is known, motion compensation can be achieved by solving a general matrix inversion problem [2]. A 'shot' here is defined as a subset of k-space, but can be extended to include the k-space data for an entire image; i.e., in dynamic single-shot or real-time acquisition. The assumption in [2] is that all k-space points within a shot are acquired during a sufficiently short period to be considered motion-free. Although the theory has been established, this methodology is rarely applied mainly due to the difficulty of estimating dense deformation fields for every shot. An attempt to bypass this problem was published in [3], where a motion model within the field of view is assumed and parameterized as a linear combination of selected input basis signals, e.g. navigator echoes or respiratory belts or signals derived from ECG. The optimization process is extended to interleave the estimation of a motion-free image and a motion model by consecutively solving two large linear systems. Although this approach avoids the estimation of deformation fields, it requires a more complicated matrix representation and couples motion estimation and compensation. Significant computational cost is thus incurred to converge to a solution.

With the aim to correct non-rigid motion, we propose an efficient algorithm as an extension of the motion compensation framework based on matrix inversion. The key feature is to estimate the pixel-wise forward and backward deformation fields using the inverse-consistent non-rigid registration algorithm, and to perform multi-coil reconstruction using eigen-mode analysis. In this way, the motion field can be interpreted as a prior, simplifying the solver and reducing total reconstruction times to the order of seconds. The algorithm was successfully applied to cardiac real-time cine imaging showing suppression of ghosting artifacts caused by chest-wall motion.

2 Methods

2.1 General Motion Compensation Framework

Suppose the MR acquisition consists of k_s shots, presenting a $n_1 \times n_2$ image. The k-space sampling pattern for every shot is defined by the sampling matrix $\boldsymbol{A_s}$ $(n_1 n_2 \times n_1 n_2)$. The motion-free image $\boldsymbol{S_0}$ $(n_1 n_2 \times 1)$ is corrupted by the motion field

$\boldsymbol{u_s}$ $(n_1 n_2 \times n_1 n_2)$. Then the final image with motion artifacts $\boldsymbol{S}$ $(n_1 n_2 \times 1)$ can be expressed by the following matrix equation, as proposed in [3]:

$$\boldsymbol{S} = \left(\sum_{s=0}^{k_s-1} \boldsymbol{F^H} \cdot \boldsymbol{A_s} \cdot \boldsymbol{F} \cdot \boldsymbol{u_s}\right) \cdot \boldsymbol{S_0} = \boldsymbol{g} \cdot \boldsymbol{S_0} \quad (1)$$

Here $\boldsymbol{F}$ $(n_1 n_2 \times n_1 n_2)$ is the Fourier transform and $\boldsymbol{F^H}$ is the inverse of $\boldsymbol{F}$ (hermitian conjugate). $\boldsymbol{u_s}$ is the image transform, corresponding to the deformation for shot s. We emphasize that $\boldsymbol{u_s} \cdot \boldsymbol{S_0}$ represents the transformed image intensities and can be efficiently computed by image interpolation operation. For every pixel in $\boldsymbol{S_0}$, its spatial position is calculated after applying the transform. The intensity at this position is estimated by performing an image interpolation. This process is repeated for all pixels in $\boldsymbol{S_0}$ and the resulting transformed image is $\boldsymbol{u_s} \cdot \boldsymbol{S_0}$. The task here is to estimate the motion-free image $\boldsymbol{S_0}$, given the measured k-space data $\boldsymbol{S}$, sampling pattern $\boldsymbol{A_s}$ and deformation field $\boldsymbol{u_s}$. While the Eq. 1 is linear, it is fully capable to represent both rigid and non-rigid motion, because only transformed image intensities, not the motion itself, are needed.

To estimate the motion-free image $\boldsymbol{S_0}$, the inversion of ghosting matrix $\boldsymbol{g}$ (here 'ghosting' means motion artifacts are introduced after applying this matrix to motion-free images) is necessary. Despite the large size of matrix $\boldsymbol{g}$, the standard conjugate gradient solver such as LSQR only requires the computation of matrix-vector product $\boldsymbol{g} \cdot \boldsymbol{S}$ and $\boldsymbol{g^H} \cdot \boldsymbol{S}$, as pointed out in [2]. These matrix-vector products can be efficiently computed using image pixel-wise operations, such as FFT and image interpolation.

Given the matrix description of motion compensation and its fast solver, the deformation field $\boldsymbol{u_s}$ for every shot is still missing. Besides, $\boldsymbol{g^H} = \sum_{s=0}^{k_s-1} \boldsymbol{u_s^H} \cdot \boldsymbol{F^H} \cdot \boldsymbol{A_s} \cdot \boldsymbol{F}$ is needed since $\boldsymbol{g^H} \cdot \boldsymbol{S}$ is required by the LSQR solver. One way to get $\boldsymbol{g^H}$ is to explicitly compute $\boldsymbol{g}$. But this will disable the fast algorithm using image based operations. Fortunately, as suggested in [2], if the inverse deformation $\boldsymbol{u_s^{-1}}$ is available in the sense of inverse consistency $\boldsymbol{u_s^{-1}} \cdot \boldsymbol{u_s} = \boldsymbol{id}$, the $\boldsymbol{u_s^H}$ can be replaced by $\boldsymbol{u_s^{-1}}$. Then, fast pixel-wise operations can be applied to computing $\boldsymbol{g} \cdot \boldsymbol{S}$ and $\boldsymbol{g^H} \cdot \boldsymbol{S}$.

2.2 Inverse-Consistent Non-rigid Registration

Incorporating motion compensation into MR reconstruction requires the availability of both forward and backward deformation fields $\boldsymbol{u_s}$ and $\boldsymbol{u_s^{-1}}$ for every shot. While most non-rigid algorithms neither supply the inverse deformation field as the output nor maintain the inverse consistency, there are some researches in this topic [4,5]. To estimate both forward and backward deformation, we propose to utilize an inverse consistent non-rigid registration algorithm [6] which estimates both deformation fields using an interleaved optimization scheme and maintaining the symmetry and inverse-consistency of image alignment. In this optimization scheme, a symmetric energy functional is descended by alternating the registration direction after each iteration and enforcing the inverse consistency. This inverse-consistency optimization is added on top of a variational registration framework [6]. Although current algorithm is selected because of its great efficiency and ability to produce the pixel-wise deformation field which is required by the matrix description of motion compensation, other inverse consistent registration algorithms can be applied.

2.3 Motion Compensated MR Reconstruction with Multiple Coils

The usage of multiple phased array receiver coils has become essential in contemporary MR systems due to the success of parallel imaging. It is thus necessary to utilize the proposed technique in the context of multi-coil imaging using parallel image acquisition.

There are two approaches that extend the above-mentioned technique for multi-coil imaging. Firstly, one could repeat the solver to correct motion independently for every coil with the same deformation fields. After all coils are corrected, they can be combined to generate the final image. To estimate the deformation fields, non-rigid registration needs to be applied to the magnitude image of every shot. There are many possible ways to achieve this. In dynamic imaging, an initial round of parallel imaging reconstruction can be applied to generating images needed to estimate the deformation fields. In multi-segment 3D imaging, images for every shot could be computed by a rough re-gridding reconstruction. In 2D imaging, those images can be obtained by a simple SENSE reconstruction. Other alternatives include performing the registration on low-resolution pre-scan before running the intended protocol.

An apparent drawback of performing the solver on every coil is the increased computational cost since coils with 32 channels or more are commonly used in clinical settings. To avoid this, suppose a total of n_r coils are used for imaging, the ghosting matrix formula can be listed for every coil: $\boldsymbol{S_r} = \left(\sum_{s=0}^{k_s-1} \boldsymbol{F^H} \cdot \boldsymbol{A_s} \cdot \boldsymbol{F} \cdot \boldsymbol{u_s}\right) \cdot \boldsymbol{C_r} \cdot \boldsymbol{S_0} = \boldsymbol{g_r} \cdot \boldsymbol{S_0}$ where $\boldsymbol{C_r}$ is the coil sensitivity for the channel r. The above matrix equation can be repeated for every coil and stacked together except $\boldsymbol{S_0}$, ending up solving a linear system of $n_r n_1 n_2 \times n_1 n_2$. The pixel-wise image operation can still be applied for efficiency. The main disadvantage here is that the coil sensitivity must be known beforehand, which can be achieved if a reference scan is feasible. However, in many applications a precise estimation of coil sensitivity is problematic due to motion of the chest-wall or abdomen that alters the coil sensitivity between the pre-scan and subsequent image acquisitions.

We propose to employ the so-called 'eigen-coil' method to reduce computational cost without affecting reconstruction accuracy. The eigen-coil images are computed by performing a Karhunen-Loeve Transform or principal component analysis on the multi-coil images [7]. Suppose a set of n_r coils, each coil acquires an image of $n_1 \cdot n_2$ pixels at the same time. These n_r images can be represented by a $n_r \times n_1 n_2$ data matrix $\boldsymbol{D}$. As the same object is imaged by all coils and there are overlaps between coil sensitivities, images from every coil bear redundancy. Therefore, the empirical covariance matrix of $\boldsymbol{D}$, defined as $\boldsymbol{DD^H/n_1 n_2}$, has maximal n_r non-zero eigenvalues. The eigen-coil images are computed by multiplying the data matrix by corresponding eigenvectors. If we sort the eigenvalues by its magnitude and the first few eigen-coil images will occupy most image information, while those corresponding to small eigenvalues are basically representing noise. Therefore, it is adequate to only perform the motion compensation on the first few eigen-coil images without discernibly jeopardizing the accuracy. In this way, the computational cost can be largely reduced and coil sensitivity is not required.

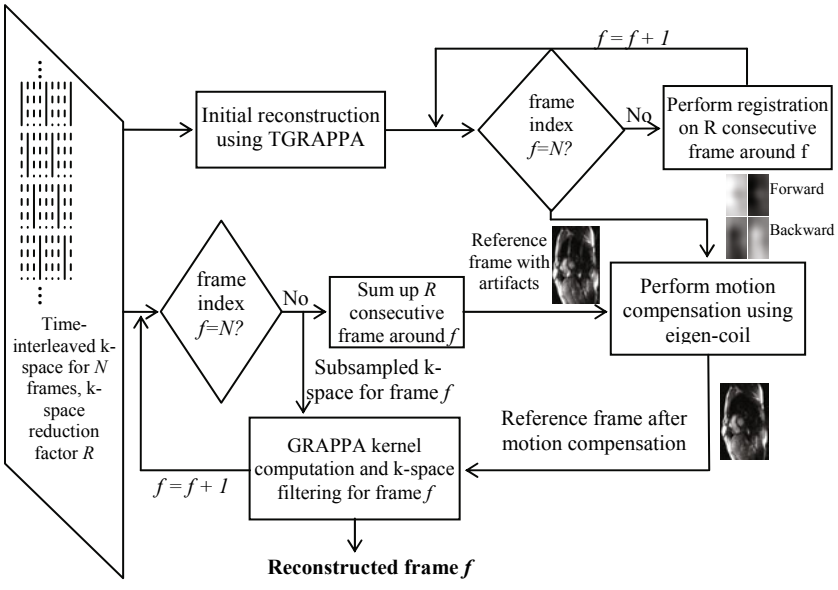


Fig. 2. Motion compensation reconstruction for time-interleaved real-time cine imaging

2.4 Motion Compensated Real-Time Cine Imaging

Real-time cine imaging is a technique that captures cardiac motion without the need for breath-holding and regular cardiac rhythm; these are requirements of segmented multi-shot acquisition strategies. Temporal GRAPPA, i.e. TGRAPPA [8], and temporal SENSE, i.e. TSENSE [9] are state-of-the-art dynamic parallel imaging methods that acquire time-interleaved, undersampled k-space data and fuse information from adjacent frames in order to estimate the coil sensitivity for TSENSE or the autocalibrating signals (ACS) for TGRAPPA. This fusion of information from k-space data acquired in an interleaved fashion can lead to ghosting artifacts in the estimated coil sensitivity that might corrupt the reconstructed image. This can be especially problematic in the real-time stress imaging, where patient heart-rate and respiratory motion are at extremes, and there may be severe mismatches between the estimated coil sensitivity and the acquired image data .We propose here to apply the presented technique to correct the chest-wall motion between adjacent frames and suppress artifacts in the coil sensitivity estimation, which leads to an improved reconstruction. Fig. 2 illustrates the workflow of this reconstruction. In this scheme, an initial reconstruction was performed using standard TGRAPPA and the frame-to-frame deformation fields were estimated from the magnitude images. To reconstruct frame f, every R neighboring k-spaces around f were treated as R shots from a MR acquisition and a motion-compensated reference image was computed by solving the general matrix inversion with the deformation fields as inputs. This reference was fed into the GRAPPA computation to reconstruct frame f.

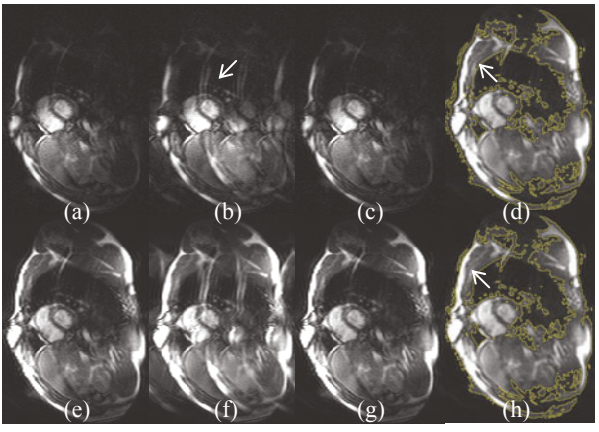


Fig. 3. Simulation of non-rigid motion compensation for multi-coil complex image. K-space was divided into 4 shots and deformation fields were applied to mimic the chest wall motion. (a) Original image of the first coil; (b) Ghosting image of first coil due to non-rigid chest wall motion. (c) After motion compensation, the image content is completely recovered. (e-g) Sum-of-square original image, its ghosting version and recovered result. (d,h) Effect of chest wall motion.

3 Results

Proposed algorithm was implemented using Matlab (MathWorks, Natick, Massachusetts) and the non-rigid registration was programmed in C++. All computations were performed on a dual-core desktop with 3.00GHz CPU and 6GB RAM without utilizing multi-threading. Typical computation time needed for non-rigid registration is less than 0.1ms (320×80 pixels). The conjugate gradient solver costs ~1s for every eigen-coil image. In all following tests, the solver was fixed to iterate 15 times due to the observation that more iterations did not result in better results.

3.1 Simulation

To test feasibility, two simulations were performed. The first was designed to apply a known non-rigid deformation to a complex 32 channel cardiac cine image. Four artificial deformation fields were applied to warp this image to simulate a continuous chest-wall motion. K-space sub-sampling was restricted to the phase-encoding direction (horizontal axis in this case) and a regular sampling pattern with 4 times reduction was used and leads to 4 shots for every coil. After contaminating these 4 shots with corresponding deformation fields, the motion compensation solver was performed on each coil independently and thus repeated 32 times. Simulation results are shown in Fig. 3, presenting an accurate correction that is virtually identical to the ground truth. The second simulation was designed to perform motion compensation on the eigen-coil images. Fig. 4 shows the motion compensation results on eigen images where we empirically selected 0.05 as the cutoff of accumulated eigenvalue and kept 10 modes out of 32. Motion correction was applied to those kept and results

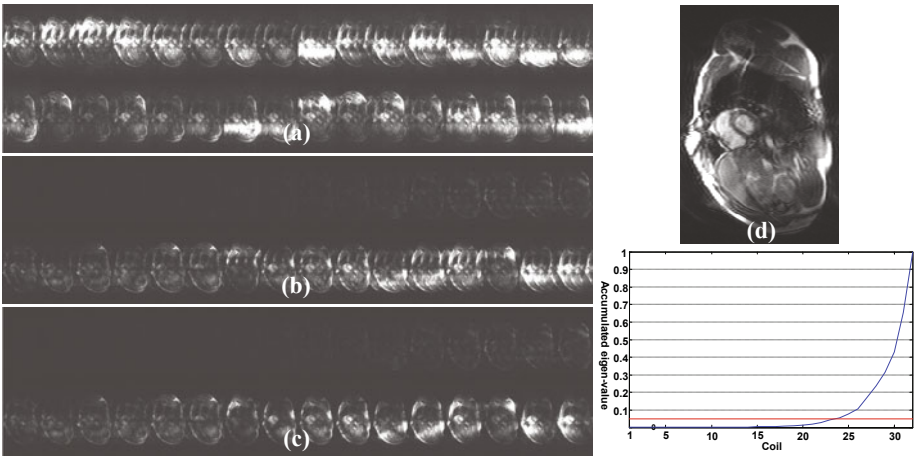


Fig. 4. Simulation of motion compensation using eigen-coil. (a) Ghosting artifacts appear on every coil; (b) All eigen modes bear ghosting artifacts, but only those corresponding to large eigenvalues need correction. (c) Eigen images after correction show good artifact removal. (d) Sum-of-square image after correcting 10 modes and leaving other 22 unchanged. Difference compared to Fig. 3(f) where all coils are processed is indiscernible with ~70% time-saving. The curve shows the accumulated eigen-values for all 32 channels.

were transformed back to original image space by multiplying the hermitian transpose of eigenvector matrix. As 95% of the total image content was actually compensated for motion, the corrected image was indiscernible compared to previous simulation with ~70% of processing time reduced.

3.2 In Vivo Study

18 volunteers (10 men, 8 women; mean age 36.7±15.2 years) underwent time-interleaved free-breathing real-time cine examinations. The study protocol was approved by the Institutional Review Board and all participants gave written informed consent. A clinical 1.5T MR scanner (MAGNETOM Avanto, Siemens) equipped with a 32 channel phased-array coil (Rapid MR International, Columbus, Ohio, USA) was used. MR sequence parameters included: balanced SSFP readout, TR=1.09/TE=0.9ms, acquired matrix 160×80 (interpolated to 160×120), flip angle 58°, interpolated in-plane resolution 2.44×2.44 mm^2, slice thickness 10mm, bandwidth 1420Hz/pixel and parallel imaging reduction factor of 4. For every patient, 9 slices were acquired to cover the left ventricle of the heart. Both standard TGRAPPA and motion compensated version were performed on all datasets. Image qualify was assessed by both noise and artifacts levels. To measure image noise, a retrospective signal-noise-ratio (SNR) estimation algorithm based on random matrix theory was applied [10]. This method has been validated for dynamic imaging. Although images reconstructed with or without motion compensation showed similar SNR (26.8 vs. 28.1), the advantage of propose method was that motion induced artifacts were clearly reduced when the chest all motion is

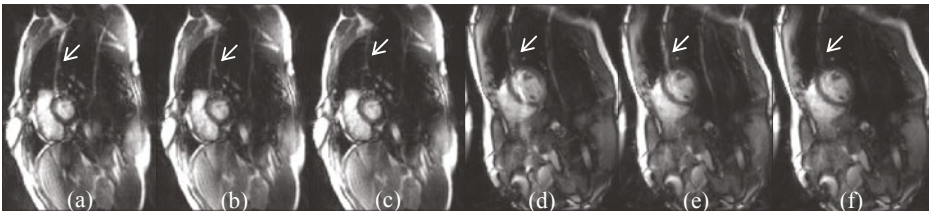


Fig. 5. Example images generated using three different methods to estimate ACS signals for TGRAPPA reconstruction. (a,d) Averaging all undersampled k-space; (b,e) Moving averaging every 4 consecutive frames; (c,f) Moving averaging every 4 frames with motion compensation, where artifacts induced by chest wall motion were better suppressed.

severe during the image acquisition, as illustrated in Fig. 5. The motion compensation via eigen-coil method was found to be robust and visual equivalent to the per-coil strategy, while the reduction of reconstruction time is around 60~70%.

4 Conclusion and Discussion

This paper describes a novel MR motion compensation algorithm based on inverse-consistent non-rigid registration and general matrix inversion. A practical realization of this method was proposed for real-time cine imaging. The proposed method requires a multi-shot interleaved k-space acquisition where each shot is free of motion inconsistency. In case this condition is violated, a longer shot can be divided into multiple shorter ones. The specific implementation for cardiac cine imaging relies on 2D image registration, implying that through-slice motion is not specifically handled. For future work, we plan to extend the proposed method to compensate multi-shot 3D imaging as the registration is not limited to 2D. Further validation studies are being pursued with emphasis on the clinical benefits of our technique.

Acknowledgments. Authors would like to thank Dr Philip Batchlor (King's College London) for discussion and providing general matrix inversion program.

References

1. Atkinson, D.: Motion Correction. Sunrise Course: Image Reconstruction. In: ISMRM (2011)
2. Batchelor, P.G., Atkinson, D., Irarrazaval, P., Hill, D., Hajnal, J., Larkman, D.: Matrix description of general motion correction applied to multishot images. MRM 54(5), 1273–1280 (2005)
3. Odille, F., Vuissoz, P., Marie, P., Felblinger, J.: Generalized Reconstruction by Inversion of Coupled Systems (GRICS) applied to free-breathing MRI. MRM 60(1), 146–157 (2008)
4. Christensen, G.E., Johnson, H.J.: Consistent image registration. TMI 20(7), 568–582 (2001)
5. Hernandez, M., Bossa, M., Olmos, S.: Registration of anatomical images using paths of diffeomorphisms parameterized with stationary vector field flows. IJCV 85, 291–306 (2009)

6. Guetter, C., Xue, H., Chefd'hotel, C., Guehring, J.: Efficient symmetric and inverse-consistent deformable registration through interleaved optimization. IEEE ISBI (2011)
7. Jolliffe, I.: Principal component analysis. Springer, New York (2002)
8. Breuer, F., Kellman, P., Griswold, M., Jakob, P.: Dynamic Autocalibrated parallel imaging using Temporal GRAPPA (TGRAPPA). MRM 53, 981–985 (2005)
9. Kellman, P., Epstein, F., McVeigh, E.: Adaptive sensitivity encoding incorporating temporal filtering (TSENSE). MRM 45(5), 846–852 (2001)
10. Ding, Y., Chung, Y., Simonetti, O.P.: A method to assess spatially variant noise in dynamic MR image series. MRM 63, 782–789 (2010)

Trans-Dimensional MCMC Methods for Fully Automatic Motion Analysis in Tagged MRI

Ihor Smal, Noemí Carranza-Herrezuelo, Stefan Klein,
Wiro Niessen, and Erik Meijering

Biomedical Imaging Group Rotterdam,
Departments of Medical Informatics and Radiology
Erasmus MC – University Medical Center Rotterdam, The Netherlands

Abstract. Tagged magnetic resonance imaging (tMRI) is a well-known noninvasive method allowing quantitative analysis of regional heart dynamics. Its clinical use has so far been limited, in part due to the lack of robustness and accuracy of existing tag tracking algorithms in dealing with low (and intrinsically time-varying) image quality. In this paper, we propose a novel probabilistic method for tag tracking, implemented by means of Bayesian particle filtering and a trans-dimensional Markov chain Monte Carlo (MCMC) approach, which efficiently combines information about the imaging process and tag appearance with prior knowledge about the heart dynamics obtained by means of non-rigid image registration. Experiments using synthetic image data (with ground truth) and real data (with expert manual annotation) from preclinical (small animal) and clinical (human) studies confirm that the proposed method yields higher consistency, accuracy, and intrinsic tag reliability assessment in comparison with other frequently used tag tracking methods.

Keywords: Tracking, motion analysis, tagged MRI, cardiac imaging, particle filtering, Markov chain Monte Carlo.

1 Introduction

In magnetic resonance imaging (MRI), tissue tagging by spatial modulation of magnetization (SPAMM) [2] has shown great potential for studying myocardial motion and treatment effects after myocardial infarction. Contrary to conventional MRI, tagged MRI (tMRI) allows for quantitative assessment of the myocardium, for example by analyzing regional strain. With tMRI, two orthogonal sets of magnetic saturation planes, each orthogonal to the image plane, can be created in short time. The deformation of the resulting tag pattern over time reflects the deformation of the underlying cardiac tissue, and is of clinical significance for assessing dynamic properties of the heart. Existing methods, based on active contours [1,5], non-rigid image registration [4], and optical flow [3], are not robust or accurate enough to deal with the strongly varying image quality of typical experimental data. The inevitable fading of the tag pattern and acquisition noise (Fig. 1) also affect commercially available tracking methods, based on harmonic phase (HARP) MRI [11].

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 573–580, 2011.

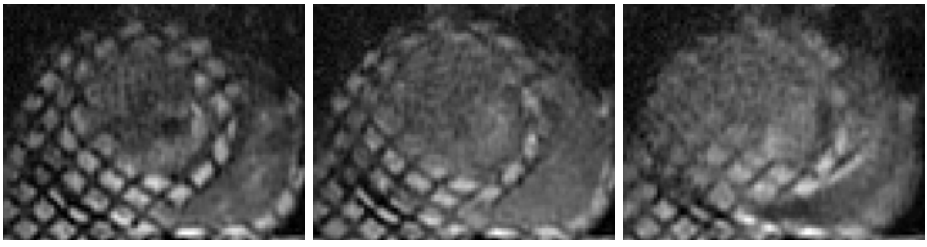


Fig. 1. Examples of images from studies of rat hearts, with frames 6, 12, and 18 (out of 24 per heart cycle) showing the fading of the tagging

In this paper we propose a novel Bayesian method for motion analysis in tMRI, implemented by means of particle filtering (PF), and a trans-dimensional Markov chain Monte Carlo (TDMCMC) approach. Using a specially designed tag likelihood measure and heart motion model (based on non-rigid image registration), the method allows for more robust and accurate tag position estimation in a sequence of noisy images. Moreover, TDMCMC makes statistical inference about both the *existence* and *position* of hundreds of appearing and disappearing tags possible and tractable. We demonstrate how tag existence can be used for automatic detection of left ventricular (LV) contours (or the myocardium). The performance of the proposed method is evaluated using realistic synthetic image data of several types, and real data from both preclinical (small animal) and clinical (human) experiments.

2 Proposed Tracking Framework

2.1 Particle Filtering and Importance Sampling

Bayesian estimation infers the posterior distribution $p(\mathbf{s}_t|\mathbf{z}_{1:t})$ of the unobserved state $\mathbf{s}_t$ of a system (the tag pattern in our case), which changes over time, taking into account noisy measurements (images in our case) $\mathbf{z}_{1:t} \triangleq \{\mathbf{z}_1, \ldots, \mathbf{z}_t\}$ up to time t. The conceptual solution is a two-step recursion $p(\mathbf{s}_t|\mathbf{z}_{1:t-1}) = \int D(\mathbf{s}_t|\mathbf{s}_{t-1})p(\mathbf{s}_{t-1}|\mathbf{z}_{1:t-1})d\mathbf{s}_{t-1}$ and $p(\mathbf{s}_t|\mathbf{z}_{1:t}) \propto L(\mathbf{z}_t|\mathbf{s}_t)p(\mathbf{s}_t|\mathbf{z}_{1:t-1})$, where two application specific models (the state transition $D(\mathbf{s}_t|\mathbf{s}_{t-1})$ and the likelihood $L(\mathbf{z}_t|\mathbf{s}_t)$) need to be specified. In practice, $p(\mathbf{s}_t|\mathbf{z}_{1:t})$ is obtained using a particle filtering (PF) approximation [6], which represents the posterior by a set of N_s random samples ("particles"), and associated normalized weights $\{\mathbf{s}_t^{(i)}, w_t^{(i)}\}_{i=1}^{N_s}$. With standard PF, these samples and weights are then propagated through time to give an approximation of the filtering distribution at subsequent time steps as $\mathbf{s}_t^{(i)} \sim D(\mathbf{s}_t|\mathbf{s}_{t-1}^{(i)})$ and $w_t^{(i)} \propto L(\mathbf{z}_t|\mathbf{s}_t^{(i)})$, $i = \{1, \ldots, N_s\}$, $t = \{1, 2, \ldots\}$. At each time step, the optimal state is estimated from $p(\mathbf{s}_t|\mathbf{z}_{1:t})$, for example using the MMSE estimator as $\hat{\mathbf{s}}_t = \sum_{i=1}^{N_s} w_t^{(i)}\mathbf{s}_t^{(i)}$.

More advanced and efficient PF variants (such as sequential importance sampling (SIS) [6]) choose the weights using an importance function describing which

areas of the state space contain most information about the posterior. Here, we propose to use as importance function a time-dependent transform, $\mathbf{T}_t$, obtained using mutual-information based non-rigid image registration (NRR) [4], to predict the position of tags. The transform maps any point (x, y) in the reference image $I_{t=0}$ to its corresponding point (x', y') in the image I_t at time t, and serves to *approximately* estimate the global motion of the myocardium, after which particle sampling can be constrained to the relevant parts of the state space, much closer to the optimal solution. To utilize $\mathbf{T}_t$ as importance function, we transform the original image sequence $\{I_t\}_{t=0}^{T-1}$ to form new measurements (images) $\mathbf{z}_t(x, y) = I_t(\mathbf{T}_t(x, y))$, which all resemble the initial frame $I_{t=0}$. The proposed PF-based approach is subsequently applied to $\mathbf{z}_t$ and refines the initial coarse tag position estimation (obtained by NRR).

2.2 Refining the Particle Filtering Framework

At each time step t, the configuration of M tags (taken as the intersections of the lines in the tagging pattern) is represented by a set of state vectors $S_t = \{\mathbf{s}_{m,t}\}_{m=1}^{M}$. Each tag m is described by $\mathbf{s}_{m,t} = (x_{m,t}, y_{m,t}, \theta_{m,t}, e_{m,t})$, where $(x_{m,t}, y_{m,t})$ and $\theta_{m,t}$ define the spatial position and local orientation, respectively. The binary "existence" variable $e_{m,t}$ indicates whether tag m lies within $(e_{m,t} = 1)$ or outside $(e_{m,t} = 0)$ the myocardium. In the initial configuration at $t = 0$, the coordinates $(x_{m,0}, y_{m,0})$ of the state vectors correspond to the intersections of perfectly orthogonal tag lines, and the local orientation $\theta_{m,0}$ is known from the imaging protocol (see also Fig. 2). For each tag m with $e_{m,t} = 1$, a neighborhood system $N_{m,t} = \{m' : m \sim m', e_{m',t} = 1\}$ is defined, where "$\sim$" indicates that tag m' belongs to one of the four possible nearest neighbors of tag m. At each time point, the set of "live" tags $E_t = \{m : e_{m,t} = 1\}$ defines the lattice $\mathbf{s}_t = \{\mathbf{s}_{m,t} : m \in E_t\}$, which corresponds to the myocardium.

Likelihood. For each tag m, the likelihood $L_s(\mathbf{z}_t|\mathbf{s}_{m,t}, N_{m,t})$ of the state $\mathbf{s}_{m,t}$ is dependent on the tag neighborhood, and given by one of 16 models (Fig. 2). The complete likelihood is factorized as $L(\mathbf{z}_t|\mathbf{s}_t) = \prod_{m \in E_t} L_s(\mathbf{z}_t|\mathbf{s}_{m,t}, N_{m,t})$. To measure the likelihood of the state $\mathbf{s}_{m,t}$, the model is positioned at $(x_{m,t}, y_{m,t})$ with orientation $\theta_{m,t}$, the mean and variance of the image intensities in the "black" and "gray" regions (Fig. 2) are computed (μ_0 and σ_0^2 versus μ_1 and σ_1^2), and the discriminative power of the t-value (related to the Student's t-test) is used to measure whether the intensity samples from the likelihood model in the two regions are from the same distribution. Thus, the likelihood is defined as $L_s(\mathbf{z}_t|\mathbf{s}_{m,t}, N_{m,t}) = \max\left\{0, (\mu_1 - \mu_0)(\sigma_1^2 n_1^{-1} + \sigma_0^2 n_0^{-1})^{-1/2}\right\}$, where n_0 and n_1 denote the number of counted samples in the corresponding regions of the model. Due to the fact that a larger number of samples is used to estimate the means and variances, the proposed likelihood model is both accurate and very robust, even for extremely low-SNR image data.

Transition Prior and Tag Interactions. The state evolution is modeled as $D(\mathbf{s}_t|\mathbf{s}_{t-1}) = D_s(\mathbf{s}_t|\mathbf{s}_{t-1})D_N(\mathbf{s}_t)$, where $D_N(\mathbf{s}_t)$ models tag interactions, and the

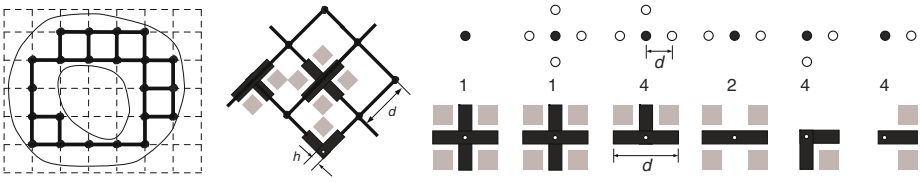


Fig. 2. From left to right: example of initial configuration with "live" (black dots) tags and the corresponding schematic LV contours interpolation (gray curves), positioning of the likelihood models depending on the neighborhood ($\theta_{exp} = 45°$), and the possible neighborhood configurations and corresponding models with indicated number of possible unique rotations (d and h are known from the acquisition protocol)

multi-tag transition prior is factorized as $D_s(\mathbf{s}_t|\mathbf{s}_{t-1}) = \prod_{m\in E_t} D_s(\mathbf{s}_{m,t}|\mathbf{s}_{m,t-1})$, which is valid due to the fact that the global (correlated) tag motion was removed by the transform $\mathbf{T}_t$. To assess the remaining (uncorrelated) deviations of the tags in $\mathbf{z}_t$ from the perfect square grid, the transition prior for $(x_{m,t}, y_{m,t}, \theta_{m,t})$ of a tag is modeled as a constrained random walk. State transitions are further constrained by explicitly modeling the interaction between the tags using a Markov random field (MRF) [9] with respect to the neighborhood systems, using a spring-like model $D_N(\mathbf{s}_t) = \prod_{m\in E_t} \prod_{m'\in N_{m,t}} \Phi(\mathbf{s}_{m,t}, \mathbf{s}_{m',t})$, where $\Phi(\mathbf{s}_{m,t}, \mathbf{s}_{m',t})$ is maximal for tags separated by distance d, and goes to zero if the distance is larger or smaller than d (see also [9]), which penalizes neighboring tag locations that are either too close or too far from each other.

Initialization. The initial tag positions for configuration S_0 are obtained by a completely automatic initialization procedure [13], which matches the position of an artificial pattern of orthogonal lines (defined by θ_{exp}, d, and h) with the observed tag pattern in the image $I_{t=0}$. Each tag is tested if it falls into the myocardial region or the background, using the likelihood L_s and the one-tailed t-test: $e_{m,0} = 1$ if the p-value < 0.01, and $e_{m,0} = 0$ otherwise.

2.3 MCMC-Based Particle Filtering

Tracking of multiple objects using PF in a joint state space (where the state vector $\mathbf{s}_t$ is a collection of individual object states) suffers from exponential complexity in the number of tracked objects, rendering it infeasible for more than a few objects [9,6]. To deal with a large and variable number of tags (variable dimensionality of the state vector $\mathbf{s}_t$), we employ TDMCMC-based PF [9]. MCMC methods define a Markov chain over the space of configurations, such that the stationary distribution of the chain is equal to the sought posterior $p(\mathbf{s}_t|\mathbf{z}_{1:t})$. The simulation of the chain (the sets of samples $\{\mathbf{s}_t^{(i)}\}_{i=1}^{N_s}$ representing the posterior), using for example the Metropolis-Hastings (MH) algorithm [7], includes proposing a new state $\mathbf{s}_t^*$ (as $\mathbf{s}^{(i)}$), sampled from a proposal density $q(\mathbf{s}_t^*|\mathbf{s}_t^{(i-1)})$, and accepting or rejecting it according to an acceptance ratio $\alpha = \min\left(1, p(\mathbf{s}_t^*|\mathbf{z}_{1:t})q(\mathbf{s}_t^{(i-1)}|\mathbf{s}_t^*)/(p(\mathbf{s}_t^{(i-1)}|\mathbf{z}_{1:t})q(\mathbf{s}_t^*|\mathbf{s}_t^{(i-1)}))\right)$.

In TDMCMC [8], the approach is modified to define a Markov chain over a variable-dimensional state space. The algorithm first selects a move type v from a finite set of allowed moves designed to explore a variable-dimensional state space. We propose three moves, which increase the dimensionality of the state (tag birth), decrease it (tag death), or leave it unchanged (tag update). Choosing the move type is done by sampling from a prior distribution $p(v)$. TDMCMC PF generates new samples from the proposal distribution $q(\mathbf{s}_t^*|\mathbf{s}_t^{(i-1)})$, where the move-specific proposal $q_v(m)$ selects (uniformly from the "live" or "dead" tags) a tag that should be updated, added, or removed from $\mathbf{s}_t$, and the tag-specific proposal $q_v(\mathbf{s}_t^*|\mathbf{s}_t, m)$ (in our case $N_s^{-1}\sum_i D_s(\mathbf{s}_t|\mathbf{s}_{t-1}^{(i)})$) samples its new state. In this case, the corresponding acceptance ratios can be readily computed [8,9].

Having defined the necessary moves, models and acceptance ratios, we run the TDMCMC PF to obtain an approximation of the filtering distribution $p(\mathbf{s}_t|\mathbf{z}_{1:t})$, where the Markov chain constructed by the TDMCMC represents the believe distribution of the current tag states given the observations. Point estimates are obtained by computing the MMSE position of each tag. The filtering distribution is also used to compute an estimate of tag existence in time at any point within the myocardium by computing the MMSE estimate $\hat{e}_{m,t} = N_s^{-1}\sum_i e_{m,t}^{(i)}$. This information is further used to segment the myocardium by averaging the existence variables for each tag $\hat{e}_{m,t}$ over time and thresholding at value 0.5. The resulting boundary tags are used for spline interpolation of the LV contours. The final result is a dense motion field estimation together with the indication of the tag existence within the myocardial region during motion. The results of tag tracking using the TDMCMC PF are used to refine the B-spline representation of the dense displacement field (initially obtained by the NRR step) using the B-spline refinement procedure proposed in [10].

3 Experimental Results

3.1 Evaluation on Synthetic Image Data

The performance of tag tracking depending on image quality (SNR) was assessed using synthetic image sequences, for which the ground truth was available. To realistically simulate LV motion through the cardiac cycle, we used a numerical phantom [3], which modeled simultaneous radial expansion/contraction and rotation of "tagged" myocardium (Fig. 3). The image sequences (30 frames of size 256×176 pixels) were created for SNRs of $\{18.06, 12.04, 8.51, 6.02\}$ dB, by corrupting the noise-free images with Rician noise (Fig. 3), which is typical in MRI. Additionally, the typically observed fading of the tag pattern during a heart cycle was modeled, resulting in sequences with time-varying SNR (changing from 18.06 dB in the initial frame to 6.02 dB in the last frame).

The accuracy of tag tracking was evaluated using the root mean square error (RMSE) in tag positions, averaged for each method over five independent runs on different realizations of the synthetic data. The NRR step was done using

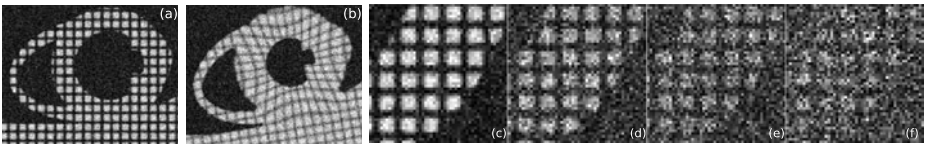


Fig. 3. Example of synthetic data with time-varying SNR (modeling tag fading) in frame 1 (a) and frame 10 (b), and zoomed image regions demonstrating the modeled tag and myocardium appearance for different SNR levels, corresponding to 18.06 dB (c), 12.04 dB (d), 8.51 dB (e), 6.02 dB (f)

Table 1. The results of tracking accuracy assessment (RMSE as a function of SNR), where the numbers represent RMSE ± standard deviation, given in pixel units

SNR	6.02 dB	8.51 dB	12.04 dB	18.06 dB	Time-varying SNR
HARP	1.99±3.15	1.15±8.56	0.51±0.66	0.39±0.88	1.17±1.79
Standard PF	0.94±0.78	0.76±0.56	0.43±0.45	0.29±0.32	0.53±0.37
TDMCMC PF	0.64±0.37	0.39±0.22	0.24±0.16	0.13±0.10	0.34±0.25
Manual	1.24±0.71	0.88±0.55	0.71±0.50	0.61±0.41	0.64±0.38

`elastix`, an open source image registration toolbox (`http://elastix.isi.uu.nl/`). The parameters for the TDMCMC simulation were fixed to $p(death) = p(birth) = 0.1$, $p(update) = 0.8$, $N_s = 100000$. All other parameters were optimized for best performance. The results of tag tracking using TDMCMC PF are shown in Table 1, in comparison with manual analysis, tracking using HARP [11] and a "standard" PF implementation (recently proposed by Smal et al. [12], which does not use NRR-based importance sampling). Our method clearly performs superiorly. The computational costs are given in Table 2.

3.2 Evaluation on Real Image Data

Experiments on real tMRI data were conducted on a clinical 3T MRI scanner (GE Medical Systems). Image sequences were collected from studies on healthy and diseased rats and pigs, and diseased human patients. Multiple short-axis view images of size 256×256 pixels were acquired with the following imaging parameters {human-, rat-, pig-related}: repetition time {6.5, 13, 4} msec, echo time {3.1, 4, 1.25} msec, flip angle {12, 7, 11} degrees, slice thickness {8, 1.6, 6} mm, spacing between slices {28, 1.6, 12} mm, pixel size $\{1.48\times1.48, 0.19\times0.19, 1.25\times1.25\}$ mm^2, frames per heart cycle {20, 24, 20}, number of slice positions {3, 7, 4}, tag spacing {7.7, 1.5, 6} mm, and $\theta_{exp} = 45°$. Five tMRI sequences of each type (human, rat, pig) were analyzed using the proposed method. The algorithm parameters were fixed to the same values as in the experiments on synthetic data. As no ground truth was available this time, we measured the accuracy in comparison with manual tracking produced by experts.

The results of the comparison are given in Table 3 and typical tracking results in comparison with HARP are shown in Fig. 4. Visual inspection reveals that the proposed method correctly estimated the tag intersections within the

Table 2. Average computational times in [s/frame] for considered image sequences

	Time
HARP	1.3
Standard PF	3.4
NRR+TDMCMC PF	11.7+10.2

Table 3. The results of tracking accuracy assessment (RMSE) using real image data from human, pig and rat experiments

	Human	Rat	Pig
HARP	1.65±1.73	1.36±0.93	1.43±1.35
Standard PF	1.48±1.43	1.35±0.87	1.27±1.17
TDMCMC PF	1.23±1.02	1.33±0.84	1.12±0.84

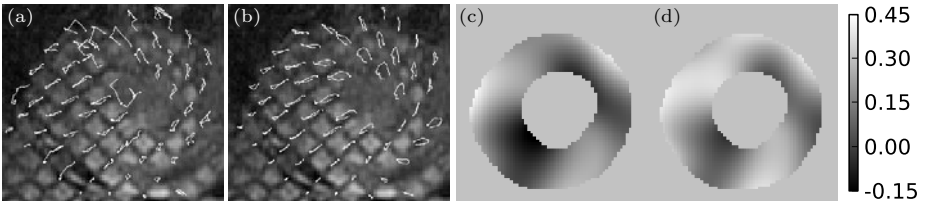


Fig. 4. Tag trajectories obtained using HARP (a) and TDMCMC PF (b), with examples of Lagrangian circumferential strain maps obtained using TDMCMC PF for frames 12 (c) and 16 (d) of the human-heart image data

myocardium through the whole image sequence. Contrary to our method, HARP produced many erroneous tracks, and reliable tag tracking was achieved only in the first 3–6 frames of the image sequence (where the tags had not faded yet) or in regions within the myocardium with sufficiently high SNR. The results in Table 3 do not include these erroneous tracks, because of their absence in the manually annotated data (it was difficult and tedious even for the experts to track the myocardial boundary points). Thus, the results in the table are biased towards tags that could be tracked relatively well, and the differences between the methods would likely be larger if all tags would be included.

4 Conclusions

Motion analysis in cardiac tMRI is a challenging problem in practice due to the poor quality of image data and complex motion scenarios. In this paper we have proposed and evaluated a novel Bayesian approach to tag tracking in tMRI, which combines prior knowledge about the heart dynamics (obtained using non-rigid image registration) with modeling of the myocardium appearance in noisy tMRI data. Straightforward generalization of the Bayesian formulation to the problem of multi-tag tracking is computationally prohibitive due to the increase in dimensionality of the state space. Therefore we have proposed a novel particle filtering (PF) method based on a trans-dimensional Markov chain Monte Carlo (TDMCMC) approach that efficiently deals with high-dimensional spaces and can track varying numbers of tags over time. The new method, which was evaluated using both synthetic and real tMRI data, demonstrated higher tracking consistency, accuracy and robustness in comparison with the commercially available HARP and a standard PF approach. Apart from yielding more accurate tracking results, the proposed TDMCMC PF is also capable of detecting

the presence/absence of tags within the myocardium in a probabilistic fashion, which can be used to signify whether the tracking results in the different parts of the myocardium are reliable or not. The analysis results (dense displacement and strain maps, obtained after refining the initial B-spline representation of the dense displacement field (obtained by the NRR step) with the tag tracking using the TDMCMC PF) can be used to develop new classification techniques for automatic diagnosis of healthy and diseased patients. This will be a subject of future work. As yet, the method is already being employed in ongoing longitudinal experiments in our institute, with the primary goal to quantify left ventricular remodeling after myocardial infarction in small animals.

References

1. Amini, A.A., Chen, Y., Curwen, R.W., Mani, V., Sun, J.: Coupled B-snake grids and constrained thin-plate splines for analysis of 2-D tissue deformations from tagged MRI. IEEE Trans. Med. Imaging 17(3), 344–356 (1998)
2. Axel, L., Dougherty, L.: MR imaging of motion with spatial modulation of magnetization. Radiology 171, 841–845 (1989)
3. Carranza-Herrezuelo, N., Bajo, A., Sroubek, F., Santamarta, C., Cristobal, G., Santos, A., Ledesma-Carbayo, M.J.: Motion estimation of tagged cardiac magnetic resonance images using variational techniques. Comput. Med. Imaging Graph. 34, 514–522 (2010)
4. Chandrashekara, R., Mohiaddin, R.H., Rueckert, D.: Analysis of 3-D myocardial motion in tagged MR images using nonrigid image registration. IEEE Trans. Med. Imaging 23, 1245–1250 (2004)
5. Deng, X., Denney Jr., T.S.: Three-dimensional myocardial strain reconstruction from tagged MRI using a cylindrical B-spline model. IEEE Trans. Med. Imaging 23(7), 861–867 (2004)
6. Doucet, A., de Freitas, N., Gordon, N.: Sequential Monte Carlo Methods in Practice. Springer, Berlin (2001)
7. Gilks, W.R., Richardson, S., Spiegelhalter, D.J.: Markov Chain Monte Carlo in Practice. Chapman and Hall, Boca Raton (1996)
8. Green, P.: Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika 82, 711–732 (1995)
9. Khan, Z., Balch, T., Dellaert, F.: MCMC-based particle filtering for tracking a variable number of interacting targets. IEEE Trans. Pattern Anal. Mach. Intell. 27, 1805–1819 (2005)
10. Lee, S., Wolberg, G., Shin, S.Y.: Scattered data interpolation with multilevel B-splines. IEEE Trans. Vis. Comp. Graph. 3, 228–244 (1997)
11. Osman, N.F., Kerwin, W.S., McVeigh, E.R., Prince, J.L.: Cardiac motion tracking using CINE harmonic phase (HARP) magnetic resonance imaging. Magn. Reson. Med. 42, 1048–1060 (1999)
12. Smal, I., Niessen, W.J., Meijering, E.: Particle filtering methods for motion analysis in tagged MRI. In: 7th IEEE International Symposium on Biomedical Imaging, pp. 488–491. IEEE Press, Piscataway (2010)
13. Urayama, S., Matsuda, T., Sugimoto, N., Mizuta, N., Yamada, N., Uyama, C.: Detailed motion analysis of the left ventricular myocardium using an MR tagging method with a dense grid. Magn. Reson. Med. 44, 73–82 (2000)

4-D Generative Model for PET/MRI Reconstruction

Stefano Pedemonte[1], Alexandre Bousse[2], Brian F. Hutton[2], Simon Arridge[1], and Sebastien Ourselin[1]

[1] The Centre for Medial Image Computing, UCL, London, United Kingdom
[2] Institute of Nuclear Medicine, UCL Hospitals NHS Trust, London, United Kingdom

Abstract. We introduce a 4-dimensional joint generative probabilistic model for estimation of activity in a PET/MRI imaging system. The model is based on a mixture of Gaussians, relating time dependent activity and MRI image intensity to a hidden static variable, allowing one to estimate jointly activity, the parameters that capture the interdependence of the two images and motion parameters. An iterative algorithm for optimisation of the model is described. Noisy simulation data, modeling 3-D patient head movements, is obtained with realistic PET and MRI simulators and with a brain phantom from the BrainWeb database. Joint estimation of activity and motion parameters within the same framework allows us to use information from the MRI images to improve the activity estimate in terms of noise and recovery.

Keywords: Molecular Imaging, Emission Tomography, Motion correction, Multi-modality, Bayesian Networks.

1 Introduction

Resolution of pharmaceutical concentration in emission tomography is limited by photon count statistics and by motion of the patient [1]. The recent development of imaging systems that combine Emission Tomography and MRI in the same machine is enabling new biological and pathological analysis tools for clinical and pre-clinical research [2]. Inherent co-registration and simultaneity of the acquisitions introduce a number of advantages over the separate modalities, including improved image fusion, motion correction and enhancement of the resolution of the functional image [3], posing new algorithmic and computational challenges. Several publications have focused on improvement of the activity estimate by means of an intra-patient anatomical image, for combined or sequential imaging systems, relying on the assumption that activity is related to the underlying anatomy, which is linked to the MRI or CT image intensity. Methods in the literature fall into three main categories: methods that favor a piecewise uniform reconstruction by segmenting the anatomical image and subsequently applying a smoothing prior within each identified region [3]; methods that explicitly extract boundary information from the anatomical image and relax the effect of a global smoothing prior across the identified edges [4] and methods

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 581–588, 2011.

based on information theoretic similarity functionals [5]. All such methods assume perfect co-registration of the two images and motion free acquisitions; however in brain imaging typical translations in the range of 5-20 mm and rotations of 1-4 deg are observed during PET and SPECT scans, concentrated in large sporadic movements occurring at discrete times [1], justifying event-driven motion compensation methods based on motion information provided by an external motion-tracking device [6]. Less successful methods to correct for patient movements involve the division of the scan into a number of frames, followed by spatial registration of the reconstructed images, relying on registration of highly photon limited image frames. In the context of combined PET/MRI we propose an event-driven motion compensation algorithm based on the acquisition of an MRI image each time large motion is detected by a tracking system or by an MRI navigator. Inspired by the Unified Segmentation [7] framework, in order to describe the interdependence of the variables in the multi-modal system we introduce a unified framework based on a 4-D joint generative model that allows us to estimate motion parameters from the MRI image and from the photon counts, obtaining a time consistent estimate of activity that accounts for its relation with the underlying anatomy, imaged by MRI.

2 Method

The model that is described is an extension to four dimensions of a bivariate Gaussian Mixture (GM) where the interdependence of the two imaging modalities is captured by assuming that image intensity y produced by the MRI sequence and activity λ are the expression of a hidden discrete variable $k = 1, 2, .., N_k$ representing anatomical/functional states (such as healthy gray matter, healthy white matter, hypoactive gray matter). k is considered the only reason of covariability of y and λ: $y \perp \lambda | k$. In order to account for a number of reasons of uncertainty of y (electronic noise, partial volume) and of λ (density of receptors, active radio-pharmaceutical molecules, perfusion), the probability distribution function (pdf) of y and λ is assumed to be a Gaussian of unknown parameters conditionally to k: $p(\lambda_b|k) = \mathcal{N}(\lambda_b, \theta_\lambda)$ and $p(y_b|k) = \mathcal{N}(y_b, \theta_y)$, where b indexes voxels and θ are the parameters of the Gaussians.

Considering the time dimension, with t indexing discrete time frames corresponding to detected motion events, motion is described by considering that the hidden states move. The hidden state is defined at the centre of voxel locations $X_b = (X_{b,1}, X_{b,2}, X_{b,3})$ in a reference space at time $t = 0$, assumed to be the instant the scan starts. Motion at time t warps the anatomical/functional state: $k^{[t]}(Y_b^{[t]}) = k(X_b)$, with $Y_b^{[t]} = T_{\gamma^{[t]}} X_b$ being the coordinates of body space warped by transformation $T_{\gamma^{[t]}}$ of parameter $\gamma^{[t]}$. As the body (the hidden state) deforms over time, at time t the MRI intensity in $Y_b^{[t]}$ is related to the hidden variable in X_b, being a realisation of the random process described by the GM of parameters $\theta_y^{[t]} = (\mu_{y_k^{[t]}}, \sigma_{y_k^{[t]}})$. For compactness $y_b^{[t]} = y^{[t]}(T_{\gamma^{[t]}} X_b)$; $\lambda_b = \lambda(X_b)$:

$$p(y_b^{[t]}|k^{[t]}(T_\gamma^{[t]} X_b)) = p(y_b^{[t]}|k(X_b)) = \mathcal{N}(y_b^{[t]}, \mu_{y_k^{[t]}}, \sigma_{y_k^{[t]}}) \quad (1)$$

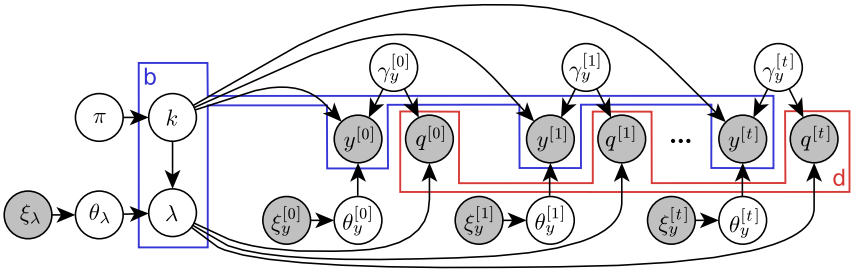


Fig. 1. Directed Acyclical Graph (DAG) of the proposed model. Observed (known) quantities are shaded. Voxels are indexed by b and lines of response (LOR) of the PET detector are indexed by d. The hidden anatomical/functional state k determines activity λ and, along with deformation parameters $\gamma^{[t]}$, determines MRI intensity $y^{[t]}$ at each time frame t. The dependence of activity and MRI intensity from the hidden state is parameterised by θ, with prior distribution of parameters ξ. Activity and deformation parameters determine photon counts $q^{[t]}$ at time t.

Activity in a voxel $Y_b^{[t]}$ at time t, assuming steady state distribution of the pharmaceutical, is a single realisation of the random process that relates it to the hidden state: $\lambda^{[1]}(T_{\gamma^{[1]}}X_b) = \lambda^{[2]}(T_{\gamma^{[2]}}X_b) = .. = \lambda(X_b)$. Photon counts $q_d^{[t]}$ along line of response (LOR) d are a realisation of the Poisson process with expected value given by the activity warped by the transformation at time t, or equivalently by activity λ_b and a time dependent transition matrix $P^{[t]} = \{p_{bd}^{[t]}\}$

$$p(\lambda_b|k) = \mathcal{N}(\lambda_b, \mu_{\lambda_k}, \sigma_{\lambda_k}) \qquad p(q^{[t]}|\lambda) = \prod_d \mathcal{P}(\sum_b p_{bd}^{[t]}\lambda_b, q_d^{[t]}) \tag{2}$$

Regarding all parameters as random variables, this model is represented by the Directed Acyclical Graph (DAG) in figure 1, where π is a multinomial distribution of the hidden state k and is here assumed to be spatially independent and unknown; ξ are hyper-parameters of the conjugate prior distribution for each of the mixture parameters. The deformation parameters $\gamma^{[t]}$ are assumed to be independent from one another conditionally to y and q as motion is considered unpredictable. Solving for the most likely state of the system provides motion parameters, activity at time $t = 0$ (and at all times when warped by $T_{\gamma^{[t]}}$) and fuzzy multi-modal segmentation $p(k_b|\lambda_b, y_b^{[t]})$ according to any combination of the MRI time frames, in fact any of $y_b^{[t]}$ (and λ) can be ignored when estimating the probability of the discrete states in a voxel. To summarise, the model accounts for the existence of an underlying tissue variable that explains the dependence of the two imaging modalities; the warping of tissue due to motion is described by a parametric deformation model. The deformation parameters, activity (in the reference space) and the parameters of the bivariate GM that describes the MR imaging system and the uncertainty about radio-pharmaceutical concentration are unknown. For a given estimate of the parameters, the marginal $p(k_b|\lambda_b, y_b^{[t]}, \theta_y, \gamma_y, \theta_\lambda)$ represents a probabilistic classification of the tissue in the reference space, according to the MRI and the projections at all time frames.

2.1 Inference

We aim at computing a point estimate of the unknown variables that maximises the joint *pdf*. Reasoning on the graph (Markov blankets), marginalising over k, all variables are independent of motion parameter $\gamma^{[t]}$ conditionally to $\lambda, \theta_y^{[t]}, \pi$; λ is independent of all other variables given $\theta_\lambda, \pi, q, \gamma$; $\theta_y^{[t]}$ is independent of all other variables conditionally to $y^{[t]}, \gamma^{[t]}, \pi$ and π is independent of all other variables given $\lambda, y, \gamma, \theta$. Starting from an initial estimate of each variable, always increasing joint *pdf* is obtained by updating the unknowns iteratively by the Iterated Conditional Modes (ICM) algorithm [8], converging to a local minimum. For compactness the expected value of the hidden state in b given λ_b, $y_b^{[t]}$ and all the parameters is defined as z_{bk}

$$z_{bk} = p(k_b|\lambda_b, y_b^{[1]}, \ldots, y_b^{[t]}) = \frac{\pi_k \mathcal{N}(\lambda_b, \mu_{\lambda_k}, \sigma_{\lambda_k}) \prod_t \mathcal{N}(y_b^{[t]}, \mu_{y_k^{[t]}}, \sigma_{y_k^{[t]}})}{\sum_{k=1}^{N_k} \pi_k \mathcal{N}(\lambda_b, \mu_{\lambda_k}, \sigma_{\lambda_k}) \prod_t \mathcal{N}(y_b^{[t]}, \mu_{y_k^{[t]}}, \sigma_{y_k^{[t]}})}$$

(i) $p(\gamma^{[t]}|y, q, \theta, \lambda, \pi)$ Given activity and the GM parameters, the deformation parameters are updated by maximising the partial probability distribution by gradient ascent. For each t: $\gamma_g^{[t](n+1)} = \gamma_g^{[t](n)} + \beta \frac{\partial \log p(\gamma^{[t]}|y,q,\theta,\lambda,\pi)}{\partial \gamma_g^{[t]}}\Big|_{\gamma^{[t](n)}}$

$$p(\gamma^{[t]}|y, q, \theta, \lambda, \pi) = \prod_b \left[\sum_k p(k) p(y_b^{[t]}|k, \theta) p(\lambda_b|k, \theta) \prod_{r \neq t} p(y_b^{[r]}|k, \theta) \right] p(q^{[t]}|\lambda, \gamma^{[t]})$$

Substituting (1) and (2) and differentiating its logarithm, by the chain rule, it simplifies to the following:

$$\frac{\partial \log p(\gamma^{[t]}|y, q, \theta, \lambda, \pi)}{\partial \gamma_g^{[t]}} = \sum_b \left[\frac{\partial y^{[t]}(T_{\gamma^{[t]}} X_b)}{\partial \gamma_g^{[t]}} \sum_k z_{bk} \frac{\mu_{y_k} - y^{[t]}(T_{\gamma^{[t]}} X_b)}{\sigma_{y_k}^2} \right] + \sum_d \left[\sum_b -p_{bd} \frac{\partial \lambda(T_{\gamma^{[t]}}^{-1} X_b)}{\partial \gamma_g^{[t]}} + q_d^{[t]} \frac{p_{bd} \frac{\partial \lambda(T_{\gamma^{[t]}}^{-1} X_b)}{\partial \gamma_g^{[t]}}}{\sum_{b'} p_{b'd} \lambda(T_{\gamma^{[t]}}^{-1} X_{b'})} \right]$$

where the gradient of the MRI image intensity with respect to the transformation parameters is, by the chain rule, the dot product of the gradient of the transformed image and the gradient of the coordinates Y_b with respect to the transformation parameters:

$$\frac{\partial y^{[t]}(T_{\gamma^{[t]}} X_b)}{\partial \gamma_g^{[t]}} = \nabla y^{[t]}(T_{\gamma^{[t]}} X_b) \cdot \left(\frac{\partial}{\partial \gamma_g^{[t]}} T_{\gamma^{[t]}} \right) X_b$$

The same derivation applies to the gradient of the activity with respect to the transformation parameters.

(ii) $p(\lambda|\gamma, \pi, q, \theta) = p(\lambda|\gamma, y, \pi, \theta) p(q|\lambda, \gamma)$ Given the deformations and the GM parameters, activity is estimated by the EM algorithm for Poisson likelihood,

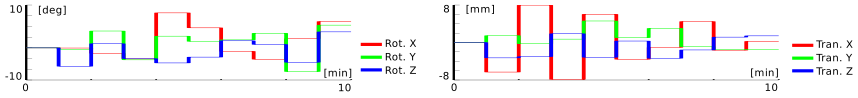


Fig. 2. Simulated random rigid motion parameters. Rotation in degrees (left) and translation im mm (right) along the three axes.

which is preferred over other gradient type optimisation algorithms as it guarantees positivity. The One Step Late approximation is adopted as the M step isn't otherwise solvable in closed form for any $p(\lambda|\gamma, y, \pi, \theta)$, as discussed in [9]:

$$\hat{\lambda}_b^{(n+1)} = \hat{\lambda}_b^{(n)} \frac{1}{\sum_{d,t} p_{bd}^{[t]} + \frac{\partial}{\partial \lambda_b} \log p(\lambda|\gamma, y, \pi, \theta)\Big|_{\lambda_b^{(n)}}} \sum_{d,t} \frac{p_{bd}^{[t]}\, q_d^{[t]}}{\sum_{b'=1}^{N_b} p_{b'd}^{[t]}\, \hat{\lambda}_{b'}^{(n)}} \tag{3}$$

$$p(\lambda|\gamma, y, \pi, \theta) = \prod_b \left[\sum_k p(k) p(\lambda_b|k, \theta) \prod_t p(y_b^{[t]}|k, \theta, \pi) \right]$$

Differentiating and expanding, the gradient in (3) simplifies to the following:

$$\frac{\partial \log p(\lambda|\gamma, y, \pi, \theta)}{\partial \lambda_b} = \sum_k z_{bk} \frac{\mu_{\lambda_k} - \lambda_b}{\sigma^2_{\lambda_k}}$$

(iii) $p(\theta_y^{[t]}|y^{[t]}, \gamma^{[t]}, \pi)$ Given activity and the transformation parameters, the parameters of the GM and the latent prior probability of the hidden state are updated by EM as it has better convergence properties than other gradient based algorithms: the expected value of the unobserved variables z_{bd} is updated using the provisional estimate of the parameters and the parameters are updated by:

$$\hat{\mu}_{\lambda_k}^{(n+1)} = \frac{1}{N_b} \frac{\sum_{b=1}^{N_b} z_{bk}^{(n+1)} \lambda_b}{\hat{\pi}_k^{(n+1)}} \qquad \hat{\sigma}^{2^{(n+1)}}_{\lambda_k} = \frac{1}{N_b} \frac{\sum_{b=1}^{N_b} z_{bk}^{(n+1)} (\hat{\mu}_{\lambda_k}^{(n+1)} - \lambda_b)^2}{\hat{\pi}_k^{(n+1)}}$$

$$\hat{\mu}_{y_k^{[t]}}^{(n+1)} = \frac{1}{N_b} \frac{\sum_{b=1}^{N_b} z_{bk}^{(n+1)} y_b^{[t]}}{\hat{\pi}_k^{(n+1)}} \qquad \hat{\sigma}^{2^{(n+1)}}_{y_k^{[t]}} = \frac{1}{N_b} \frac{\sum_{b=1}^{N_b} z_{bk}^{(n+1)} (\hat{\mu}_{y_k^{[t]}}^{(n+1)} - y_b^{[t]})^2}{\hat{\pi}_k^{(n+1)}}$$

$$\hat{\pi}_k = \frac{1}{N_b} \sum_b z_{bk}$$

3 Validation Study

Noisy simulation data modeling 3-D patient head movements were constructed by projecting at various orientations a brain phantom based on the BrainWeb database [11] and by warping the corresponding MRI image. Rigid motion was simulated by randomly generating rotation and translation along the three axes, according to amplitudes and frequency that resemble typical measurements of

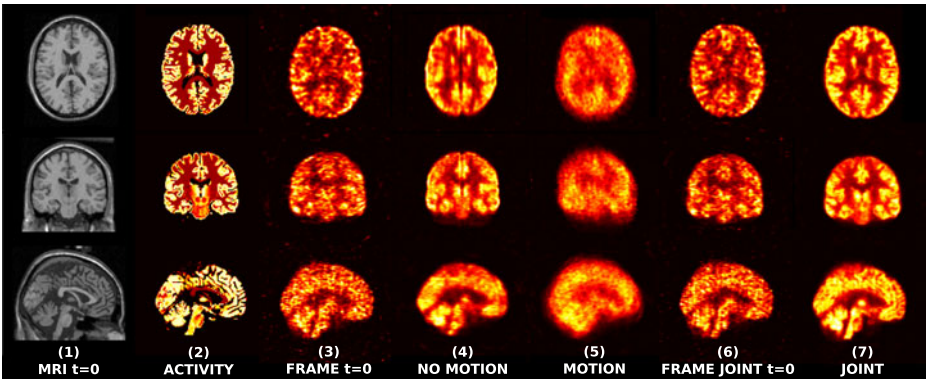


Fig. 3. From left to right: (1) MRI image frame at time $t = 0$; (2) activity phantom; (3) MLEM reconstruction of single frame ($t = 0$); (4) MLEM reconstruction of motion-free simulation; (5) MLEM reconstruction with no motion compensation; (6) reconstruction of single frame using joint generative model; (7) joint estimation of motion and activity. The method estimates correctly the rigid motion parameters, and reduces noise.

head movement within a scanner for Emission Tomography [1] (figure 2). The MRI and functional imaging processes were decoupled by running independent simulations based on the ground truth normal brain tissue model from BrainWeb. The MRI image was generated with the BrainWeb simulator, which realistically accounts for noise of the imaging system. The parameters of the simulator were set for T1-weighted imaging with noise standard deviation set at 3% of the brightest tissue and perfect uniformity of the magnetic field (in accordance with the simplistic GM model). Activity of $^{18}F\text{-}FDG$ was simulated by associating typical activity levels to different tissue types, proportionally to partial voxel occupation. Specifically the activity in gray matter was set to a value 4 times higher than in all other tissues. The total number of counts was set to 32 Million. A $1.9M$ LOR PET imaging system was simulated by means of a rotation-based projector with realistic position dependent point source response [10] and ideal pre-reconstruction correction for randoms, scatter and attenuation. The MRI and activity images were defined on a cubic grid of $(128 \times 128 \times 128)$ voxels. The number of tissue types was assumed to be $N_k = 4$; for each t, $\mu_{y_k^{[t]}}$ were initialised to evenly spaced values in the range of intensity of the MRI image; $\sigma_{y_k^{[t]}}$ were initialised to $1/N_k$ of the image intensity range; μ_{λ_k} were initialised to evenly spaced values between 0 and the maximum activity assigned to the phantom; σ_{λ_k} to $1/N_k$ of the maximum activity assigned to the phantom; the mixing coefficients to $\pi_k = 1/N_k \; \forall \; k \in N_k$. Though ICM can be applied in any order, in this validation study the transformations and the parameters of the GM were updated for all time points at the same time and activity was estimated considering projections at all time frames, repeating 5 iterations of (i), (ii), (iii) for 20 iterations (early termination). Results of the reconstruction are reported in figures 3 and 4. The activity estimate produced by applying the joint generative model presents higher coefficients of recovery both in gray matter and

	GM RC	GM SNR	WM RC	WM SNR
SINGLE FRAME	0.71	2.4	0.76	2.2
NO MOTION	0.71	4.0	0.76	3.0
MOTION	0.68	2.9	0.60	3.5
JOINT FRAME	0.72	2.7	0.77	2.3
JOINT	0.78	4.3	0.82	3.2

Fig. 4. Left: multi-modal anatomical/functional classification ($p(k_b|\lambda_b, y_b^{[t]})$). Right: Recovery Coefficient (RC) and Signal to Noise Ratio (SNR) in the gray matter and white matter regions. The table refers to the images in figure 3. The proposed method improves recovery and reduces noise by virtue of the additional information provided by the MRI image.

white matter when compared to standard MLEM reconstruction of the motion affected data. Even when compared to motion free MLEM reconstruction, the activity obtained by the generative model presents slightly higher recovery and signal to noise ratio, due to the MRI images improving the activity estimate.

3.1 Computational Complexity

Computing the registration, reconstruction and segmentation jointly is costly. The proposed algorithm is essentially based on the recursive computation of projections and back-projections, exponentials (Gaussian likelihoods), spatial gradients, linear interpolations and 3-D transformations. Projections and back-projections constitute by far the most demanding part. Computing and storing the projection matrix $P = \{p_{bd}\}$ is not feasible. However the matrix is sparse and can be applied efficiently to an activity vector by a series of convolutions and summations if activity is defined on a regularly spaced grid [10]. If $P = \{p_{bd}\}$ is the probability that an event emitted in X_b in the reference space is detected in LOR d, then the expected photon counts at time t is $\bar{n}^{[t]} = \Delta t P \lambda^{[t]}(X_b)$, where Δt is the length of the time frame

$$\lambda^{[t]}(X_b) = \lambda^{[t]}(T^{-1}TX_b) = \lambda(T^{-1}X_b) \qquad \bar{n}^{[t]} = P\Delta t\lambda(T^{-1}X_b)$$

Projection and backprojection in (3) are then computed efficiently by convolution on a ragular grid as in [10] by warping the current activity estimate by each transformation, projecting, back-projecting and transforming back. This applies to any invertible transformation. A fast GPU based algorithm was used [10] along with an ordered subset scheme that updates activity considering only subsets of the lines of response in space and time, achieving convergence in 2 h.

4 Discussion

In this paper we presented a novel unified framework to reconstruct brain PET images taking into account potential motion during the acquisition and prior information coming from a second modality. With the growing field of PET/MRI scanners, we believe that such paradigm can have a tremendous impact on the

quality of the reconstructed PET images. The model has been evaluated with synthetic data, proving feasibility of the method; it is difficult, however, to envisage at this stage implementation strategies on real imaging systems such as defining the most appropriate MRI sequence and frequency of the frames. Given the general acceptance of probabilistic atlas based segmentation algorithm [7] [12], we believe that a joint generative model for PET/MRI might prove useful when coupled with population based spatially varying priors of the hidden states and hyperpriors for the parameters of the mixture model in order to include experience in the reconstruction process. Since the presented model is valid for any invertible transformation, the method could be applied to other parts of the body by using a non-rigid transformation. Application to continuously moving structures however would require continuous MR acquisition, generating a tremendous amount of data and requiring a conspicuous amount of computational resources. Possible extensions of the model may include accounting for farmacokinetics and modeling lesions as outliers.

References

1. Fulton, R.R., Meikle, S.R., Eberl, S., Pfeiffer, J., Constable, R.T., Fulham, M.J.: Correction for head movements in positron emission tomography using an optical motion-tracking system. IEEE Trans. Nucl. Sci. 49(1), 116–123 (2002)
2. Cherry, S.R.: Multimodality Imaging: Beyond PET/CT and SPECT/CT. Semin. Nucl. Med. 39(5), 348–353 (2009)
3. Atre, A., Vunckx, K., Baete, K., Reilhac, A., Nuyts, J.: Evaluation of different MRI-based anatomical priors for PET brain imaging. In: IEEE Nucl. Sci. Sym. Conf., Orlando, pp. 1–7 (October 2009)
4. Leahy, R., Yan, X.: Incorporation of Anatomical MR Data for Improved Functional Imaging with PET. In: Inf. Proc. in Med. Imag., pp. 105–120. Springer, Heidelberg (1991)
5. Somayajula, S., Rangarajan, A., Leahy, R.M.: PET image reconstruction using anatomical information through mutual information based priors: a scale space approach. In: Int. Sym. on Biomed. Imag., pp. 165–168 (April 2007)
6. Rahmim, A., Dinelle, K., Cheng, J.C., Shilov, M.A., Segars, W.P., Lidstone, S.C., Blinder, S., Rousset, O.G., Vajihollahi, H., Tsui, B.M.W., Wong, D.F., Sossi, V.: Accurate Event-Driven Motion Compensation in High-Resolution PET Incorporating Scattered and Random Events. IEEE Trans. Med. Imag. 27(8) (August 2008)
7. Ashburner, J., Friston, K.: Unified segmentation. Neuroimage 26(3), 839–851 (2005)
8. Besag, J.: On the Statistical Analysis of Dirty Pictures. J. of the Royal Stat. Soc. Series B (Methodological) 48(3), 259–302 (1986)
9. Green, P.G.: Bayesian Reconstructions From Emission Tomography Data Using a Modified EM Algorithm. IEEE Trans. on Med. Imag. 9(1), 84–93 (1990)
10. Pedemonte, S., Bousse, A., Erlandsson, K., Modat, M., Arridge, S., Hutton, B.F., Ourselin, S.: GPU accelerated rotation-based emission tomography reconstruction. In: IEEE Nucl. Sci. Sym. Conf., M13-287, pp. 2657–2661 (2010)
11. Brain Web, http://mouldy.bic.mni.mcgill.ca/brainweb/
12. Menze, B.H., van Leemput, K., Lashkari, D., Weber, M.-A., Ayache, N., Golland, P.: A Generative Model for Brain Tumor Segmentation in Multi-Modal Images. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6362, pp. 151–159. Springer, Heidelberg (2010)

Model-Based Segmentation and Motion Analysis of the Thoracic Aorta from 4D ECG-Gated CTA Images

Andreas Biesdorf[1], Stefan Wörz[1], Tobias Müller[2,3], Tim Frederik Weber[2,3], Tobias Heye[2], Waldemar Hosch[2], Hendrik von Tengg-Kobligk[2,3], and Karl Rohr[1]

[1] University of Heidelberg, BIOQUANT, IPMB, and DKFZ Heidelberg Dept. Bioinformatics and Functional Genomics, Biomedical Computer Vision Group
[2] University Hospital Heidelberg, Dept. of Diagnostic and Interventional Radiology
[3] German Cancer Research Center (DKFZ) Heidelberg, Dept. of Radiology

Abstract. Pathologies of the thoracic aorta can alter the shape and motion pattern of the aorta throughout the cardiac cycle. For diagnosis and therapy planning, determination of the aortic shape and motion is important. We introduce a new approach for segmentation and motion analysis of the thoracic aorta from 4D ECG-CTA images, which combines spatial and temporal tracking, motion determination by intensity-based matching, and 3D fitting of vessel models. The approach has been successfully applied to 30 clinically relevant 4D CTA image sequences. We have also performed a quantitative evaluation of the segmentation accuracy.

1 Introduction

The aorta is the largest vessel of the human body and exhibits a complex 3D motion pattern which is altered by cardiovascular diseases. For diagnosis and therapy planning, motion quantification of the aorta is important. However, while 3D high-resolution and time-resolved imaging of the aorta by electro-cardiogram-gated computed tomography angiography (ECG-CTA) has now become widely available, diagnosis and therapy planning in standard clinical routine still relies on static 3D images. On the other hand, manual determination of the aortic motion is difficult, time-consuming, as well as subjective, and infeasible for a large number of clinical cases. Therefore, automated methods are indispensable.

In previous work, only few approaches have been introduced for analyzing the aorta in dynamic ECG-CTA images (e.g., [5,7]). Most approaches for dynamic images have considered other imaging modalities such as MR images (e.g., [4,3,2,9]) or other vessel types such as the carotid or coronary arteries (e.g., [1,8]). Previous approaches for segmentation of the aorta from dynamic images of different modalities employ deformable models (e.g., [4,5,7]), graph-based approaches (e.g., [9]), Markov random fields (e.g. [2]), or registration-based approaches (e.g., [3]). However, in these approaches, the change of the position of the vessel centerline and the change in vessel diameters are not quantified. Moreover, while both the motion along and orthogonal to the vessel centerline are

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 589–596, 2011.

clinically important, none of the previous approaches quantifies the motion along the vessel centerline (longitudinal motion). In addition, most approaches have been developed only for 2D images (e.g., [3]), or perform vessel segmentation based on 2D slices of 3D image data (e.g., [5]). Furthermore, the segmentation of the aorta is typically computed independently for each time step (e.g., [2,5]) and time-coherence is rarely exploited (e.g., [9,7]). In [9], a 4D graph-based approach was introduced, where information over time is used for contour-based segmentation of MR images. However, changes of the vessel centerline position, the local vessel diameters, and the longitudinal vessel motion are not quantified. In [7], time-coherence is exploited for segmentation by using a biomechanical model of vessels. However, the longitudinal vessel motion cannot be determined since the model is kept fixed in this direction by boundary conditions. Also, the centerline position and the vessel diameters are not quantified.

In this contribution, we introduce a new approach for segmentation and motion analysis of the thoracic aorta from 4D ECG-CTA images. Our approach combines spatial and temporal tracking, intensity-based matching, and 3D fitting of intensity vessel models. To improve the computational efficiency and robustness, we employ a multiresolution framework. In contrast to previous approaches, we quantify both the motion of the vessel centerline and the change of local vessel diameters. Also, we explicitly determine both the *orthogonal* and the *longitudinal* vessel motion. Moreover, we exploit the time-coherence of the motion by temporal tracking based on a Kalman filter to improve the initialization and thus the computational efficiency and the accuracy. The approach has been successfully applied to 30 clinically relevant 4D CTA image sequences. We have also performed a quantitative evaluation of the segmentation accuracy.

2 Model-Based Vessel Segmentation and Motion Analysis

Our approach for the segmentation of the aortia from dynamic 4D ECG-CTA images consists of (1) model-based initialization and segmentation of the aortic arch, (2) temporal tracking based on a Kalman filter and a 3D vessel model, as well as (3) longitudinal motion determination using intensity-based matching.

2.1 3D Model-Based Initialization

For initialization of our approach, we compute a coarse segmentation of the aorta. To this end, we use a 3D parametric intensity model of vessels $g_{Cyl}(\mathbf{x}, \mathcal{R}, \sigma)$, which is a cylindrical approximation of the local vessel shape and represents the image intensities for each voxel position $\mathbf{x} = (x, y, z)^T$ within a 3D region-of-interest (ROI). The model includes parameters for the width R of the vessel and the image blur σ, and is well-suited to describe the plateau-like intensity structure of the aorta. The complete model includes intensity levels a_0 (background) and a_1 (vessel) as well as a 3D rigid transform $\mathcal{R}$ with rotation $\boldsymbol{\alpha} = (\alpha, \beta, \gamma)^T$ and translation $\mathbf{x}_0 = (x_0, y_0, z_0)^T$, which yields

$$g_{M,Cyl}(\mathbf{x}, \mathbf{p}_{Cyl}) = a_0 + (a_1 - a_0)\ g_{Cyl}(\mathcal{R}(\mathbf{x}, \boldsymbol{\alpha}, \mathbf{x}_0), R, \sigma) \tag{1}$$

with 10 parameters $\mathbf{p}_{Cyl} = (R, a_0, a_1, \sigma, \alpha, \beta, \gamma, x_0, y_0, z_0)^T$ [6].

To obtain an initial segmentation of the vessel, we use an incremental fitting procedure along the aorta that starts with a user-supplied start point, the approximate vessel orientation, as well as an estimation of the local vessel radius. The orientation is estimated based on two points along the vessel centerline and the radius is determined manually using a scalable sphere. In each increment k of the fitting procedure, the parameters of the cylinder segment are estimated by fitting the model $g_{M,Cyl}(\mathbf{x}, \mathbf{p}_{Cyl})$ to the image intensities $g(\mathbf{x})$ within a 3D ROI.

To improve the computational efficiency, we determine the initial segmentation based on the first image of a sequence with reduced isotropic resolution of 1.0mm/voxel (original resolution: 0.29-0.82mm/voxel). This improves the computational efficiency by a factor of about 10-30 and also the robustness. Note that the final segmentation result for all time points is determined using the original image resolution so that the overall segmentation accuracy is not affected.

2.2 Temporal Tracking and Determination of the Orthogonal Motion

As a result of the initialization we obtain a dense set of 200-400 vessel segments describing the geometry of the aorta in a 4D CTA image at the first time point. To determine the motion of the aorta, we compute the segmentation result for a vessel segment at each time point based on the results of the same segment from previous time points. However, note that in contrast to the previous initialization step, we here use the original image resolution and a different vessel model.

The vessel model used in the initialization has *circular* cross-sections. However, minimum and maximum vessel diameters are important in clinical applications. Thus, for the measurement using the original image resolution, we use a novel model with *elliptical* cross-sections. The vessel model represents an ideal sharp *3D elliptic cylinder* convolved with a 3D Gaussian. In contrast to the model in (1), two parameters for the minimum and maximum radius of the tubular structure, R_x and R_y, are used, and the model can be stated as

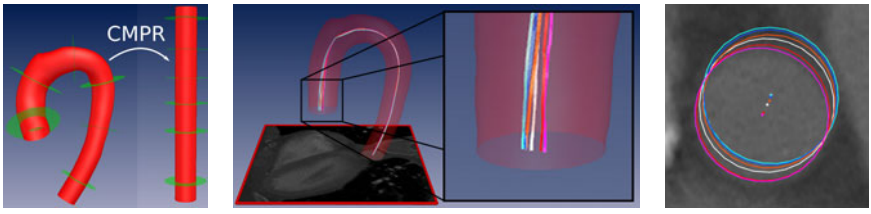
$$g_{Ell}(\mathbf{x}, R_x, R_y, \sigma) = \Phi\left(\frac{\sqrt{R_x \cdot R_y}}{\sigma} \cdot \left(1 - \sqrt{\frac{x^2}{R_x^2} + \frac{y^2}{R_y^2}}\right)\right) \tag{2}$$

where $\Phi(x) = \int_{-\infty}^{x} (2\pi)^{-1/2} e^{-\xi^2/2} d\xi$, leading to the complete model

$$g_{M,Ell}(\mathbf{x}, \mathbf{p}_{Ell}) = a_0 + (a_1 - a_0)\; g_{Ell}(\mathcal{R}(\mathbf{x}, \boldsymbol{\alpha}, \mathbf{x}_0), R_x, R_y, \sigma) \tag{3}$$

with 11 parameters $\mathbf{p}_{Ell} = (R_x, R_y, a_0, a_1, \sigma, \alpha, \beta, \gamma, x_0, y_0, z_0)^T$.

To initialize the elliptical model, we make use of the observation that the aortic arch exhibits a directed (non-random) motion throughout the cardiac cycle while there exist only relatively small deformations between subsequent time points. Therefore, we predict the vessel centerline position $\mathbf{x}_0$ for a vessel segment based on the previous centerline positions of the same vessel segment using a linear Kalman filter. The prediction $\mathbf{x}_t^*$ of the centerline position for time point $t \geq 2$ is computed by $\mathbf{x}_t^* = \boldsymbol{\Phi}_{t,t-1} \hat{\mathbf{x}}_{t-1}$, where $\boldsymbol{\Phi}_{t,t-1}$ denotes the transition matrix and



(a) (b) (c)

Fig. 1. (a) Sketch of curved multi-planar reformation. (b) Segmentation result for five time points of a 4D ECG-CTA image: Centerline and zoomed section (different time points indicated by different colors), and (c) cross-section of the ascending aorta.

$\hat{\mathbf{x}}_{t-1}$ is the estimate for the previous time point $t-1$. The estimate $\hat{\mathbf{x}}_t$ for time point t can be computed based on

$$\hat{\mathbf{x}}_t = \mathbf{x}_t^* + \hat{\mathbf{K}}_t(\mathbf{z}_t - \mathbf{H}_t\mathbf{x}_t^*), \tag{4}$$

where $\hat{\mathbf{K}}_t$ is the gain matrix, $\mathbf{z}_t$ is the measurement vector (representing the estimates from the minimization described below), and $\mathbf{H}_t$ is the measurement matrix. We use a linear motion model, i.e., $\mathbf{x}_t^* = \mathbf{x}_{t-1} + \dot{\mathbf{x}}_{t-1}\Delta t, \Delta t = 1$. The remaining parameters in $\mathbf{p}_{Ell}$ are initialized using the previous segmentation result at time point $t-1$ since we expect similar values at subsequent time points.

To estimate the parameters for one vessel segment, the 3D model $g_{M,Ell}(\mathbf{x}, \mathbf{p}_{Ell})$ is fitted to the image intensities within a 3D ROI by minimizing $\sum_{\mathbf{x}\in ROI}\left(g_{M,Ell}(\mathbf{x}, \mathbf{p}_{Ell}) - g(\mathbf{x})\right)^2$ using the method of Levenberg-Marquardt. To increase the robustness, we perform model fitting within a 3D ROI with orientation $\boldsymbol{\alpha}_{0,k} = (\alpha_{0,k}, \beta_{0,k}, \gamma_{0,k})$ of vessel segment k in the first image. Using a 3D ROI with the same orientation as the vessel centerline (but allowing rotation within the orthogonal cross-section) has the advantage that the longitudinal vessel motion can be fixed during model fitting. Note that if the longitudinal motion is not fixed, the model may slide along the vessel during fitting (which is a general problem with tubular structures since the shape in longitudinal direction hardly changes), and this could affect the spatial order of the vessel segments.

2.3 Determination of the Longitudinal Vessel Motion

The 3D motion of the thoracic aorta from 4D ECG-CTA images is difficult to determine since it is composed of different motion types including (1) motion orthogonal to the vessel centerline, (2) deformations of the aorta, as well as (3) longitudinal motion along the vessel centerline. Note that our approach for temporal tracking described above (Sect. 2.2) can determine the motion orthogonal to the vessel centerline and the deformations of the aorta (i.e., changes of the diameters). However, the longitudinal vessel motion cannot be determined.

To determine the *longitudinal* motion of the aorta, our idea is to decouple the longitudinal motion from the overall 3D motion. To this end we use curved

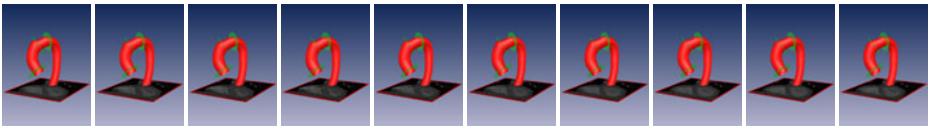


Fig. 2. Segmentation result of the aorta for a 4D ECG-CTA image at 10 time points

multi-planar reformations (CMPRs) of the aorta, a normalization of the vessel diameter for all vessel segments at a time point, and intensity-based matching. To obtain a CMPR, we resample the 3D CTA image along the vessel centerline (Fig. 1a). For each time point, we extract the sequence of orthogonal 2D slices to the vessel centerline with centers $\mathbf{x}_{0,k}$ and orientations $\boldsymbol{\alpha}_k$. For $\mathbf{x}_{0,k}$ and $\boldsymbol{\alpha}_k$, we use the estimates from temporal tracking (see Sect. 2.2 above). As a result, the curved vessel is transformed to a straight vessel. Doing this, and considering corresponding cross-sections, we can fix the *orthogonal* vessel motion. The vessel diameters, however, still can change over time which may bias the motion estimation. Therefore, we normalize the vessel radii based on the largest mean vessel radius $\overline{r}_{max}$ of all vessel segments at all time points. Doing this, we preserve the anisotropic shape and improve the robustness of intensity-based matching compared to normalization to a circle since more image information is exploited. For normalization, we perform isotropic scaling of each orthogonal 2D slice k by a factor of $s_k = \overline{r}_k / \overline{r}_{max}$, where $\overline{r}_k$ is the mean vessel radius of segment k. As a result, we obtain an image of a straight vessel with normalized vessel radii where the centerline is parallel to the z-axis. Based on the normalized CMPR, we determine the longitudinal vessel motion for each slice z_t at time point $t \geq 2$ by intensity-based matching of 3D ROIs (size: $4 \cdot \overline{r}_{max} \times 4 \cdot \overline{r}_{max} \times 5$ voxels) along the centerline in temporally successive CMPRs by minimizing

$$z_t = \arg\min \sum_{\mathbf{x} \in ROI} \left(g_t^{CMPR}(\mathbf{x}, z_t) - g_{t-1}^{CMPR}(\mathbf{x}, z_{t-1}) \right)^2 . \tag{5}$$

With this approach, the above-mentioned sliding problem for model-image matching is less severe since we perform image-image matching and exploit more information from surrounding structures (e.g., branches).

3 Experimental Results

We have applied our approach to 30 clinically relevant 4D ECG-CTA images of the thoracic aorta. The images were acquired on a SOMATOM Definition CT scanner and comprise between 168 and 423 slices with a spacing of 0.6-1.5mm and an in-plane resolution of 0.29-0.82mm. 28 4D images in our set were acquired in 5% ECG-intervals, yielding 21 3D CTA images over time for each 4D image, while two 4D images were acquired in 10% ECG-intervals, yielding 11 3D CTA images each. The images included a variety of pathologies such as atherosclerosis, aneurysms, or aortocoronary bypasses.

3.1 Quantitative Evaluation Based on Ground Truth Segmentations

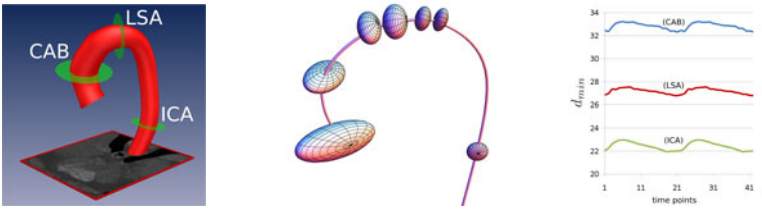
We have quantified the accuracy of our approach for determining the vessel centerline position and the minimum and maximum vessel diameter based on 5 4D ECG-CTA images with 21 time points each. Ground truth was determined independently by two radiologists for three anatomical landmarks along the aorta: (1) a coronary artery bypass branching from the ascending aorta (CAB), (2) the left subclavian artery (LSA), and (3) an intercostal artery branching from the descending thoracic aorta (ICA). The three landmarks were selected by the radiologists since they were clearly identifiable and representative to capture the range of motion in different anatomical regions (ascending aorta, aortic arch, and descending thoracic aorta, see Fig. 3a). At each landmark, the radiologists determined an orthogonal slice through the aorta and measured the centerline position as well as the minimum and maximum vessel diameter. In total, 630 manual measurements were performed ($5 \times 21 \times 3 \times 2$). For the evaluation of our approach, we used the positions of these landmarks at the first time point and automatically determined the centerline position and the minimum and maximum diameters for all subsequent time points.

To quantify the segmentation accuracy, we have computed the error of the centerline position $e_{\mathbf{x}_0}$. Since we are particularly interested in the longitudinal and orthogonal aortic motion, we also determined both components of the error, $e_{\mathbf{x}_0,long}$ and $e_{\mathbf{x}_0,orth}$. The errors are defined as the Euclidean distance between the manually obtained positions and the automatically determined positions. In addition, we have computed the errors for the minimum and maximum diameters, $e_{d_{min}}$ and $e_{d_{max}}$, defined as the absolute difference between the manual and automatic measurement. Tab. 1 gives mean values for the different measures over all images, time points, and observers.

From Tab. 1 it can be seen that for the centerline position $\mathbf{x}_0$, we obtain the best result for the landmark ICA with $\overline{e}_{\mathbf{x}_0} = 1.20\,\text{mm}$. Considering all landmarks, we obtain a mean of $\overline{e}_{\mathbf{x}_0} = 1.53\,\text{mm}$. It also turns out that $\overline{e}_{\mathbf{x}_0,orth}$ is significantly larger than $\overline{e}_{\mathbf{x}_0,long}$ (for the means we obtain $\overline{e}_{\mathbf{x}_0,long} = 0.33\,\text{mm}$ and $\overline{e}_{\mathbf{x}_0,orth} = 1.40\,\text{mm}$). This is due to the fact that the absolute motion is much larger in orthogonal than in longitudinal direction (e.g., the maximum motion in orthogonal direction is 23.7 mm and in longitudinal direction 7.2 mm). Note that for $\overline{e}_{\mathbf{x}_0,long}$ we obtain submillimeter accuracy. For $\overline{e}_{d_{min}}$ and $\overline{e}_{d_{max}}$ we also

Table 1. Mean errors for the vessel centerline position $\overline{e}_{\mathbf{x}_0}$, as well as longitudinal and orthogonal component, $\overline{e}_{\mathbf{x}_0,long}$ and $\overline{e}_{\mathbf{x}_0,orth}$, respectively, for three different landmarks. The table also includes the errors for the vessel diameters $\overline{e}_{d_{min}}$ and $\overline{e}_{d_{max}}$.

Mean error [mm] / Landmark	Centerline position			Vessel diameters	
	$\overline{e}_{\mathbf{x}_0}$	$\overline{e}_{\mathbf{x}_0,long}$	$\overline{e}_{\mathbf{x}_0,orth}$	$\overline{e}_{d_{min}}$	$\overline{e}_{d_{max}}$
(1) CAB (Ascending Aorta)	1.47	0.24	1.38	0.63	0.46
(2) LSA (Aortic Arch)	1.93	0.50	1.73	0.74	0.89
(3) ICA (Descending Aorta)	1.20	0.26	1.09	0.65	0.87
Overall Mean	**1.53**	**0.33**	**1.40**	**0.67**	**0.74**



(a) (b) (c)

Fig. 3. (a) 3D segmentation result for one time point of a 4D CTA image. The planes indicate the three landmarks (CAB, LSA, ICA, from left to right). (b) Motion represented by ellipsoids at seven points. (c) Change in d_{min} for the three landmarks.

obtain submillimeter accuracy and both values are similar. The computation time for one time point using the original image resolution on a 2.40 GHz Intel Core 2 Quad CPU is about 30 minutes.

3.2 Application to 30 Clinically Relevant 4D ECG-CTA Images

We have successfully applied our approach to 30 clinically relevant 4D ECG-CTA images of the thoracic aorta and determined the change of the vessel centerline position and the change in local vessel diameters. Fig. 1 shows the segmentation results for the centerline (b) at five time points in 3D and (c) in an orthogonal view. Fig. 2 displays the segmentation result for a 4D image at 10 time points.

Fig. 3b shows the centerline motion represented by ellipsoids for all 30 datasets at seven centerline points along the aorta. The motion ellipsoids have been determined as follows. For each of the seven centerline points and each 4D image we determine the mean of the time-varying positions. Then we align the means for all 4D images (for one centerline point) by 3D translations yielding a point cloud of 610 points ($28 \times 21 + 2 \times 11$ positions). For each point cloud, we perform a principal components analysis to determine the main axes of the ellipsoid. The size of the ellipsoid is determined using a 90% quantile. In Fig. 3b, for a better visibility, we have scaled the ellipsoids by a factor of 5. Based on the orientation and size of the ellipsoids, it can be seen that the aortic motion changes in orientation and amplitude along the aorta. While the amplitude of the motion is high near the heart (left side), the amplitude decreases with increasing distance from the heart. The lowest amplitude is at the descending aorta (right side). It can also be seen that the aorta expresses a highly twisted vessel motion.

Fig. 3c shows the change in vessel diameters for the three landmarks CAB, LSA, ICA in Tab. 1 (blue, red, and green curves) using the minimum diameter d_{min} over two cardiac cycles. As expected, d_{min} decreases with increasing distance from the heart (the curves for CAB, LSA, ICA lie above each other). Also, it can be seen that d_{min} changes periodically over the cardiac cycle. At systole (beginning of the cardiac cycle), the vessel is dilated by the increase in blood flow before it relaxes slowly back to the initial diameter.

4 Discussion

We have introduced a new approach for segmentation and motion analysis of the aorta from 4D ECG-CTA images. Our approach combines spatial and temporal tracking, motion determination by intensity-based matching, and 3D fitting of vessel models. The approach has been successfully applied to 30 clinically relevant 4D CTA images. Moreover, we have performed a quantitative evaluation of the accuracy for determining the centerline position and the diameters.

Acknowledgement. Support of the Deutsche Forschungsgemeinschaft (DFG) within the project QuantVessel (RO 247/6) is gratefully acknowledged.

References

1. Hameeteman, K., Rozie, S., Metz, C., Klein, S., van Walsum, T., van der Lugt, A., Niessen, W.: Automated carotid artery distensibility measurements from CTA using nonrigid registration. In: IEEE International Symposium on Biomedical Imaging, pp. 13–16 (2010)
2. Jodoin, P., Lalande, A., Voisin, Y., Bouchot, O., Steinmetz, E.: Markovian method for 2D, 3D and 4D segmentation of MRI. In: IEEE International Conference on Image Processing, pp. 3012–3015 (2008)
3. Lorenzo-Valdés, M., Sanchez-Ortiz, G.I., Bogren, H., Mohiaddin, R.H., Rueckert, D.: Determination of Aortic Distensibility Using Non-rigid Registration of Cine MR Images. In: Barillot, C., Haynor, D.R., Hellier, P. (eds.) MICCAI 2004. LNCS, vol. 3216, pp. 754–762. Springer, Heidelberg (2004)
4. Rueckert, D., Burger, P., Forbat, S., Mohiaddin, R., Yang, G.Z.: Automatic tracking of the aorta in cardiovascular MR images using deformable models. IEEE Transactions on Medical Imaging 16(5), 581–590 (1997)
5. Schwartz, E., Gottardi, R., Holfeld, J., Loewe, C., Czerny, M., Langs, G.: Evaluating deformation patterns of the thoracic aorta in gated CTA sequences. In: IEEE International Symposium on Biomedical Imaging, pp. 21–24 (2010)
6. Wörz, S., von Tengg-Kobligk, H., Henninger, V., Rengier, F., Schumacher, H., Bckler, D., Kauczor, H.U., Rohr, K.: 3D Quantification of the Aortic Arch Morphology in 3D CTA Data for Endovascular Aortic Repair. IEEE Transactions on Biomedical Engineering 57(10), 2359–2368 (2010)
7. Xiong, G., Taylor, C.A.: Physics-based modeling of aortic wall motion from ECG-gated 4D computed tomography. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6361, pp. 426–434. Springer, Heidelberg (2010)
8. Zhang, D.P., Risser, L., Friman, O., Metz, C., Neefjes, L., Mollet, N., Niessen, W., Rueckert, D.: Nonrigid registration and template matching for coronary motion modeling from 4D CTA. In: Fischer, B., Dawant, B.M., Lorenz, C. (eds.) WBIR 2010. LNCS, vol. 6204, pp. 210–221. Springer, Heidelberg (2010)
9. Zhao, F., Zhang, H., Wahle, A., Thomas, M., Stolpen, A., Scholz, T., Sonka, M.: Congenital aortic disease: 4D magnetic resonance segmentation and quantitative analysis. Medical Image Analysis 13(3), 483–493 (2009)

Keep Breathing! Common Motion Helps Multi-modal Mapping

V. De Luca[1,⋆], H. Grabner[1], L. Petrusca[2], R. Salomir[2], G. Székely[1], and C. Tanner[1]

[1] Computer Vision Laboratory, ETH Zürich, 8092 Zürich, Switzerland
[2] Radiology Department, Geneva University Hospital, Geneva, Switzerland

Abstract. We propose an unconventional approach for transferring of information between multi-modal images. It exploits the temporal commonality of multi-modal images acquired from the same organ during free-breathing. Strikingly there is no need for capturing the same region by the modalities. The method is based on extracting a low-dimensional description of the image sequences, selecting the common cause signal (breathing) for both modalities and finding the most similar sub-sequences for predicting image feature location. The approach was evaluated for 3 volunteers on sequences of 2D MRI and 2D US images of the liver acquired at different locations. Simultaneous acquisition of these images allowed for quantitative evaluation (predicted versus ground truth MRI feature locations). The best performance was achieved with signal extraction by slow feature analysis resulting in an average error of 2.6 mm (4.2 mm) for sequences acquired at the same (a different) time.

Keywords: motion prediction, multi-modal, ultrasound, magnetic resonance images, tracking, liver.

1 Introduction

The attention paid to organ motion due to breathing during radiation therapy has risen in recent years [4]. As treatment devices become more sophisticated, their guidance requires higher accuracy. Yet motion prediction is complicated due to the large variability in respiratory patterns, invalidating the assumption of periodic motion [4]. Unfortunately, observation of the tumor motion in real-time during therapy is often impossible, making surrogate measures of this motion essential. Common surrogates include measurements from a breathing bellow, a spirometer, optical imaging of the chest wall and imaging of the diaphragm position. The latter could be achieved by fluoroscopy, ultrasound or cine MRI, with ultrasound being the preferred choice to avoid radiation and high costs. The question remains how well the surrogates can be related to the tumor motion. Studies have shown that assuming a linear relationship between the abdominal and tumor displacements can be misleading (correlation range $[0.39 - 0.99]$) [4].

⋆ We acknowledge the Swiss National Science Foundation - Project CRSII2 127549.

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 597–604, 2011.

Statistical motion models [8,15] aim to overcome this uncertainty by learning the motion pattern for the whole organ. Tumor motion is then predicted from the motion of some landmarks in the organ. This requires that spatial correspondence between the surrogate and the organ motion model can be established, which is essentially a multi-modal image registration problem. Relating the US structures back to the pre-therapeutic MRI is not an easy task, especially in 2D. A main challenge is to define an effective image similarity measure. For liver US images, approaches include gradient images [9], matching of extracted vessels [13,11], simulations of US from CT images [17,5] and hybrids [6]. Assuming the same breathing phase, alignments are often optimized with respect to a rigid transformation. The reported registration errors were on average in the range of 3.7 mm. Additionally, tracking of US features are neither without error (e.g. 1.5 mm [2,3] for 3D US). Processing time is another issue in real-time tracking.

For a clinical setup, we propose to acquire a short pre-therapeutic 4D MRI, and record US images of the same organ during treatment. The organ under investigation is then the same before and during therapy, while the position of the image acquisition and modality have changed, see Fig. 1. Instead of multi-modal image registration of two static images, we propose a novel approach. It is based on the observation, that the appearance changes of the organ in time are mainly due to a common cause, i.e. the breathing. Inspired by recent investigation of learning invariants in time series [18], we look for low-dimensional embeddings of the image sequences in order to extract such invariant signals, which are associated to a common cause. Invariance in this context refers to independence of multi-modal image appearance and exact slice location. Summarizing, our hypothesis is that due to the presence of a global temporal pattern in the data, it is possible to relate multi-modal images of a moving organ without the need for explicit image registration.

2 Material

US and MR images of the liver were simultaneously acquired for 3 volunteers during free-breathing at the Radiology Department of Geneva University Hospital [12]. This was accomplished by modifying the US equipment to be MR compatible. The US transducer and cables were EM shielded, coupling material was added and the transducer finally mounted on a fixation frame. The sequence of 2D US images obtained from real time US 2nd-harmonic imaging (center frequency = 2.2 MHz) had a temporal resolution of 25 fps and a spatial resolution of 0.6 mm. The acquisition sequence of MRIs consisted of alternating between 2D navigator slices (at a fixed spatial position) and 2D data slices (at different locations to cover the liver) with a frequency of 2.45 Hz for the first two volunteers and 2.86 Hz for the third volunteer. 4D MR images can be created by sorting the data slices based on the similarity of the enclosing navigator slices [14]. This allows us to process 2D navigators and extend our results to 3D MR volumes. MR navigator images had a spatial resolution of 2.34 mm, 2.42 mm and 1.72 mm for the first, second and third volunteer respectively.

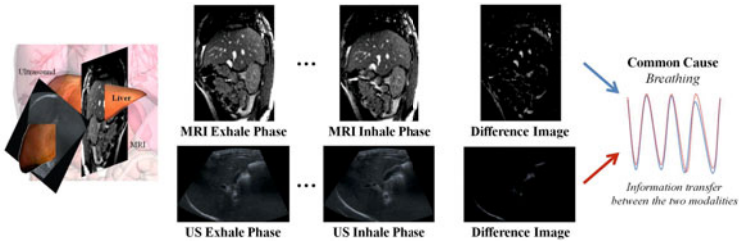


Fig. 1. Left: MR and US images of the liver are acquired at different spatial locations (e.g. sagittal MRI, right subcostal oblique US). **Center:** typical liver images from simultaneously acquired 2D MR (top) US (bottom) in exhale and inhale position are depicted. **Right:** We aim for extracting, from the image sequences, a description of the breathing as common cause, invariant of the image modality and slice location, in order to transfer image information between the two modalities.

3 Method

In this work we propose a new approach for exploiting the temporal commonality of multi-modal images (US, MRI) acquired from the liver during free-breathing. We assume that the breathing signal is statistically similar across different image modalities that represent the same organ. Hence, we extract the breathing signal from each modality, look for the similarity between these signals, and use the resulting match to predict the corresponding images. In particular, we predicted MR navigators from the observation of US images. Our method (illustrated in Fig. 2 and described in more details in Sects. 3.1-3.3) consists of 3 steps, namely

Dimensionality Reduction. Given an image sequence $\boldsymbol{I}^m$ from modality $m \in \{MRI, US\}$, we embed $\boldsymbol{I}^m$ into a low-dimensional data representation $\boldsymbol{S}^m$ using dimensionality reduction techniques.

Selection of Common Cause Signals. We select components from $\boldsymbol{S}^{MRI}$ and $\boldsymbol{S}^{US}$ characterized by a global repetitive pattern related to breathing.

Prediction of Feature Location. The prediction of image feature locations in one modality is achieved by evaluating the similarity between the breathing signals and selecting the associated images.

3.1 Dimensionality Reduction

We aim to find an invariant description of the breathing signal with respect to the image modality and independently from an exact spatial correspondence. We assume that the breathing signal lies in a low-dimension manifold extracted from the data. Numerous methods are available for dimensionality reduction [7]. However it is often unclear which method is applicable for a specific problem. We explored the following dimensionality reduction methods: Principle Component Analysis (PCA), Slow Feature Analysis (SFA) [18], Isomaps [16] and Laplacian eigenmaps [1]. Especially, SFA was proposed to capture the invariant temporal

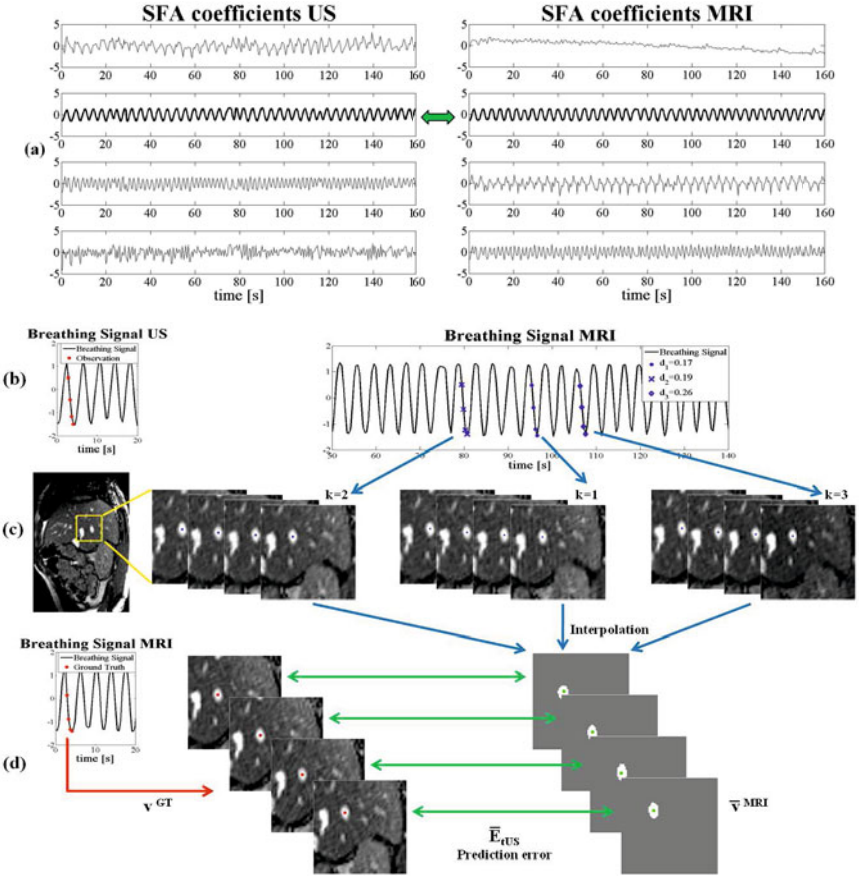


Fig. 2. (a) Low-dimensional representation of the images and selection of the breathing signals ($\hat{\boldsymbol{s}}^{US}$, $\hat{\boldsymbol{s}}^{MRI}$). **(b)** From the observation of a short sequence of $\hat{\boldsymbol{s}}^{US}$, we find the most similar sequences in $\hat{\boldsymbol{s}}^{MRI}$. **(c)** We select the associated MRIs and extract image information (vessel landmarks). **(d)** For evaluation, we interpolate the vessel landmarks ($\bar{\mathbf{v}}^{MRI}$) and calculate the prediction error $\bar{E}_{tUS}$ w.r.t. ground truth ($\mathbf{v}^{GT}$).

structures in the data by extracting slowly to quickly changing components [18]. For each modality $m \in \{MRI, US\}$, we acquire images characterized by D^m pixels at a frequency rate of f^m. We consider a time sequence of L^m images $\boldsymbol{I}^m(\tau_i^m)$, with $\tau_i^m = \tau_0^m + i/f^m$ for $0 \leq i \leq L^m - 1$. Dimensionality reduction methods transform the initial dataset $\boldsymbol{I}^m(\tau_i^m) \in \mathbb{R}^{D^m}$ into a new dataset $\boldsymbol{S}^m(\tau_i^m) = [\boldsymbol{s}_1^m(\tau_i^m); \ldots; \boldsymbol{s}_d^m(\tau_i^m)] \in \mathbb{R}^d$, with $d \ll D^m$. We used the dimensionality reduction toolbox from Laurens van der Maaten [7] for manifold learning of Isomaps and the Laplacian eigenmaps, and Matlab for PCA and SFA.

PCA. Each image $\boldsymbol{I}^m(\tau_i^m)$ is reshaped into a single column vector $\boldsymbol{x}^m(\tau_i^m) \in \mathbb{R}^{D^m}$. We calculate the mean with respect to time ($\bar{\boldsymbol{x}}^m$) and the covariance

matrix $C^{\hat{\boldsymbol{X}}} = \hat{\boldsymbol{X}}^m \hat{\boldsymbol{X}}^m$, where the i-th column of $\hat{\boldsymbol{X}}$ is equal to $\boldsymbol{x}^m(\tau_i^m) - \bar{\boldsymbol{x}}$. We then solve $\boldsymbol{C}^{\hat{\boldsymbol{X}}} \boldsymbol{w}_j^{PCA} = \lambda_j^{PCA} \boldsymbol{C}^{\hat{\boldsymbol{X}}} \boldsymbol{w}_j^{PCA}$, $\forall j \in [1, \ldots, L^m]$. λ_j^{PCA} is the i-th largest eigenvalue and $\boldsymbol{w}_j^{PCA}$ the corresponding eigenvector. In each j-th eigendirection we calculate the data projection $\boldsymbol{p}_j^{PCA} = (\boldsymbol{w}_j^{PCA})^T \hat{\boldsymbol{x}}^m$. We consider d projections so that $\boldsymbol{s}_j^m(\tau_i^m) = \boldsymbol{p}_j^{PCA}(\tau_i^m)$ for $j \in [1, \ldots, d]$.

SFA. We perform SFA on the PCA projections in order to extract signals with increasing temporal frequency from $\boldsymbol{P} = \left[\boldsymbol{p}_1^{PCA}; \ldots; \boldsymbol{p}_J^{PCA}\right] \in \mathbb{R}^{J \times L^m}$, with $J \ll L^m$. Let $\hat{\boldsymbol{P}}$ be the zero-mean data matrix. We solve the generalized eigenproblem $\dot{\boldsymbol{C}}^{\hat{\boldsymbol{P}}} \boldsymbol{w}_j^{SFA} = \lambda_j^{SFA} \boldsymbol{C}^{\hat{\boldsymbol{P}}} \boldsymbol{w}_j^{SFA}$, where $\boldsymbol{C}^{\hat{\boldsymbol{P}}}$ and $\dot{\boldsymbol{C}}^{\hat{\boldsymbol{P}}}$ are the covariance matrix of the dataset and the temporal difference, respectively. The slowest components in $\hat{\boldsymbol{P}}$ are the projections onto the eigenvectors $\boldsymbol{w}_j^{SFA}$ associated with the smallest λ_j^{SFA} and $\boldsymbol{s}_j^m(\tau_i^m) = (\boldsymbol{w}_j^{SFA})^T \hat{\boldsymbol{P}}^m$, $\forall j \in [1, \ldots, d]$.

The individual components $\boldsymbol{s}_i^m(\tau_i^m)$ are then normalized to zero mean and a standard deviation of one. The signals are further normalized to a common sampling frequency $f = min\left\{f^{MRI}, f^{US}\right\}$ (in our case $f = f^{MRI}$). We denote the normalized low-dimensional descriptions as $\hat{\boldsymbol{S}}^m(t) = [\hat{\boldsymbol{s}}_1^m(t); \ldots; \hat{\boldsymbol{s}}_d^m(t)]$, with $t = \tau_0^m + i/f$ and $0 \leq i \leq L^m - 1$. Fig. 2(a) shows examples of $\hat{\boldsymbol{S}}^{US}$ and $\hat{\boldsymbol{S}}^{MRI}$.

3.2 Selection of Common Cause Signals

We aim to select corresponding individual components in $\hat{\boldsymbol{S}}^{US}$ and $\hat{\boldsymbol{S}}^{MRI}$. Relying on the statistical assumption that a similar breathing pattern is observed in the two modalities, we want to find the components which indicate the breathing as common cause. Hence, we employ frequency analysis (i.e. Fast Fourier Transform) to the 1D eigenmodes of both modalities. Then we choose the signals characterized by a power spectral density maximum at a frequency in the range of 0.15-0.40 Hz (common breathing frequency window). We denote the selected breathing signals as $\hat{\boldsymbol{s}}^m(t)$ for modality $m \in \{US, MRI\}$.

3.3 Prediction of Feature Location

After normalization and selection of the breathing signals, we look for their similarity. We select the most similar short sequences in $\hat{\boldsymbol{s}}^{MRI}$ from a finite number of observations in $\hat{\boldsymbol{s}}^{US}$, in order to predict MR navigators. Our observation consists of a short sequence $[\hat{s}^{US}(t - (N-1)/f), \ldots, \hat{s}^{US}(t/f)]$ from the selected signal $\hat{\boldsymbol{s}}^{US}$, in order to take into account the breathing history (e.g. exhalation, inhalation). We calculate the K-nearest neighbors (K-NN) in $\hat{\boldsymbol{s}}_{MRI}$ to this sequence. In detail, we find t_k^{MRI} so that, for a given t^{US}, we minimize the distance

$$d_k = \sqrt{\sum_{n=0}^{N-1} \left[\hat{s}^{US}(t^{US} - n/f) - \hat{s}^{MRI}(t_k^{MRI} - n/f)\right]^2}, \tag{1}$$

$\forall\ 1 \leq k \leq K$, see Fig. 2(b). We considered $N = 4$ and $K = 3$. Having found the K-NN MR short sequences from the US sequence, we selected the associated

MR navigator images, see Fig. 2(c). We evaluated this approach on vessel center locations on the MR navigators. Specifically, vessels with cross-sectional cuts were semi-automatically segmented from MR navigator images[1] and their center locations $\mathbf{v}(t)$ extracted. For each of the N frames, we linearly interpolate the K center locations, to get the prediction of the vessel center position (see Fig 2(d)):

$$\bar{\mathbf{v}}^{MRI}(t^{US} - n/f) = \sum_{k=1}^{K} \left(1 - \frac{d_k}{\sum_k d_k}\right) \mathbf{v}^{MRI}(t_k^{MRI} - n/f). \tag{2}$$

Prediction Evaluation. We used data from 3 volunteers (Vol.1 - Vol.3). For each volunteer we considered one US sequence and two MRI sequences and differentiate between two setups. For *Seq1* the US and MR sequences were acquired simultaneously. The MRIs of this sequence define our ground truth (GT) data, allowing us to validate the method. For *Seq2* we use MRIs acquired minutes later and the ultrasounds from *Seq1*. This setup was used to evaluate the performance for an independent MR set, similarly to the targeted clinical application.

We also assessed the effect of replacing $\hat{\boldsymbol{s}}^{US}$ with the main diaphragm motion (first PCA eigenmode of a point displacement at the diaphragm). The diaphragm displacement was obtained from intensity-based tracking of US liver images. The tracking method was optimizing the parameters of an affine transformation of a region enclosing the diaphragm with respect to normalized cross correlation.

We quantified the prediction error by computing the mean of the Euclidean distance between the ground truth vessel center location $\mathbf{v}^{GT}$ and the corresponding location on the predicted MRIs $\bar{\mathbf{v}}^{MRI}$, see Fig. 2(d):

$$\bar{E}_{t^{US}} = \frac{1}{N} \sum_{n=0}^{N-1} \left\| \mathbf{v}^{GT}(t^{US} - n/f) - \bar{\mathbf{v}}^{MRI}(t^{US} - n/f) \right\|. \tag{3}$$

We summarized the results by the mean and standard deviation of $\bar{E}_{t^{US}}\ \forall t^{US}$.

4 Results

Table 1 lists the mean prediction error $\bar{E}$ obtained from each volunteer and tested method. It can be observed that SFA achieved on average lowest errors closely followed by PCA. Average errors increased by 1 mm (0.6 mm) for diaphragm tracking for *Seq1* (*Seq2*). When assuming no respiratory motion (using the mean position of $\mathbf{v}^{GT}$ over *Seq1* as prediction), we obtained an average prediction error over all 3 volunteers of 5.94 ± 2.11 mm (7.49 ± 3.55 mm) for *Seq1* (*Seq2*). For the best approach (SFA), we evaluated the linear dependence between the selected breathing signals $\hat{\boldsymbol{s}}^{US}$ and $\hat{\boldsymbol{s}}^{MRI}$, and between $\hat{\boldsymbol{s}}^{MRI}$ and $\mathbf{v}^{GT}$. We compared the correlation with that of the common prediction approach, i.e. diaphragm tracking to $\mathbf{v}^{GT}$, see Table 2. SFA provides on average higher correlations, supporting the better performance of SFA over diaphragm tracking for *Seq1* (Table 1).

[1] Sagittal navigator images were acquired. Hence out-of-plane motion is expected to be minor (2 mm) compared to the captured in-plane motion (15 and 8 mm) [10].

Table 1. Mean ± SD of the prediction error (in mm), see Equation (3). Best results are in bold face and second best are underlined.

		common approach	our approach			
		Tracking	Isomap	Lap.Eigenmap	PCA	SFA
Seq1	Vol. 1	3.91 ± 1.94	2.86 ± 1.66	3.73 ± 1.13	1.89 ± 0.98	**1.72 ± 0.93**
	Vol. 2	3.00 ± 1.97	4.36 ± 2.52	3.08 ± 1.76	2.98 ± 1.72	**2.89 ± 1.63**
	Vol. 3	4.02 ± 2.85	4.64 ± 3.45	3.39 ± 2.45	3.57 ± 2.45	**3.35 ± 2.37**
	Mean	3.65 ± 2.25	3.96 ± 2.54	3.40 ± 1.78	2.81 ± 1.72	**2.65 ± 1.65**
Seq2	Vol. 1	5.87 ± 3.60	5.14 ± 2.38	6.05 ± 2.59	3.58 ± 1.55	**3.36 ± 1.50**
	Vol. 2	3.99 ± 1.33	**3.31 ± 1.73**	6.40 ± 3.41	4.84 ± 2.64	4.60 ± 2.57
	Vol. 3	4.55 ± 2.69	6.98 ± 5.07	10.83 ± 6.77	**4.32 ± 2.64**	4.67 ± 2.88
	Mean	4.80 ± 2.54	5.14 ± 2.73	7.76 ± 4.26	4.25 ± 2.28	**4.21 ± 2.31**

Table 2. Correlation between extracted signals and v^{GT}

		Vol. 1	Vol. 2	Vol. 3	**Mean**
	Tracking - $\mathbf{v}^{GT}$	30.7%	84.4%	85.1%	66.7%
SFA	$\hat{\boldsymbol{s}}^{US}$ - $\hat{\boldsymbol{s}}^{MRI}$	92.3%	79.6%	91.7%	87.9%
	$\hat{\boldsymbol{s}}^{MRI}$ - $\mathbf{v}^{GT}$	87.6%	94.8%	87.6%	90.0%

5 Conclusion

We proposed an unconventional method for relating multi-modal images sequences. It is based on representing the acquired data in a low-dimensional embedding, extracting common cause signals (e.g. breathing) from both image sequences, finding the most similar sub-sequences of these signals, and using the associated images in order to predict the location of image features. The method generates the prediction in a completely unsupervised manner. Using this approach, we predicted the location of anatomical landmarks by relying on a temporal pre-therapeutical MRI sequence based on the observation of US images acquired during therapy with an average accuracy of 4.2 mm. Our performance is comparable to state-of-the-art methods (3.7 mm for multi-modal registration [11,17,6], 1.5 mm for US tracking [2,3]) while being less complex. Lowest mean errors were achieved when employing SFA. This demonstrates the advantage of explicitly using the temporal information stored in the data and supports previous investigations [18]. After these encouraging results, we plan to apply the method to 3D landmark prediction by using 4D MR techniques [14], to acquire longer image sequences and to investigate potential improvements when combining the presented approach with statistical breathing models [15].

References

1. Belkin, M., Niyogi, P.: Laplacian eigenmaps for dimensionality reduction and data rapresentation. NC 15(6), 1373–1396 (2003)
2. Foroughi, P., Abolmaesumi, P., Hashtrudi-Zaad, K.: Intra-subject elastic registration of 3D ultrasound images. Med. Image Anal. 10(5), 713–725 (2006)
3. Harris, E., Miller, N., Bamber, J., Symonds-Tayler, J., Evans, P.: Speckle tracking in a phantom and feature-based tracking in liver in the presence of respiratory motion using 4D ultrasound. PMB 55, 3363 (2010)
4. Keall, P., Mageras, G., Balter, J., Emery, R., Forster, K., Jiang, S., Kapatoes, J., Low, D., Murphy, M., Murray, B., Ramsey, C., Van Herk, M., Vedam, S.S., Wong, J., Yorke, E.: The management of respiratory motion in radiation oncology report of AAPM Task Group 76. Med. Phys. 33, 3874 (2006)
5. King, A., Rhode, K., Ma, Y., Yao, C., Jansen, C., Razavi, R., Penny, G.: Registering preprocedure volumetric images with intraprocedure 3-d ultrasound using an ultrasound imaging model. IEEE Trans. on Med. Imaging 29(3), 924–937 (2010)
6. Lange, T., Papenberg, N., Heldmann, S., Modersitzki, J., Fischer, B., Schlag, P.: 3D ultrasound-CT registration of the liver using combined landmark-intensity information. IJCARS 4(1), 79–88 (2009)
7. van der Maaten, L., Postma, E., van den Herik, H.: Dimensionality reduction: A comparative review. TiCC TR (2009-005)
8. McClelland, J., Huges, S., Modat, M., Qureshi, A., Ahmad, S., Landau, D., Ourselin, S., Hawkes, D.: Inter-fraction variations in respiratory motion models. PMB 56, 251 (2011)
9. Milko, S., Malvaer, E., Samset, E., Kadir, T.: Evaluation of bivariate correlation ratio similarity metric for rigid registration of US/MR images of the liver. IJCARS 4(2), 147–155 (2009)
10. Nguyen T., et al.: Adapting population liver motion models for individualized online image-guided therapy. In: Proc EMBS, pp. 3945–3948 (2008)
11. Penney, G., Blackall, J., Hamady, M., Sabharwal, T., Adam, A., Hawkes, D.: Registration of freehand 3D ultrasound and magnetic resonance liver images. Med. Image Anal. 8(1), 81–91 (2004)
12. Petrusca, L., De Luca, V., Arnold, P., Cecalin, Z., Goget, T., Auboiroux, V., Viallon, M., Santini, F., Terraz, S., Scheffler, K., Tanner, C., Cattin, P., Salomir, R.: Ultrasound/MR hybrid imaging: truly simultaneous motion monitoring in the abdomen and image co-registration. In: ISMRM (2011)
13. Porter, B., Rubens, D., Strang, J., Smith, J., Totterman, S., Parker, K.: Three-dimensional registration and fusion of ultrasound and MRI using major vessels as fiducial markers. IEEE Trans. on Med. Imaging 20(4), 354–359 (2002)
14. Siebenthal, M., Székely, G., Gamper, U., Boesiger, P., Lomax, A., Cattin, P.: 4D MR imaging of respiratory organ motion and its variability. PMB 52, 1547 (2007)
15. von Siebenthal, M., Székely, G., Lomax, A., Cattin, P.C.: Inter-subject modelling of liver deformation during radiation therapy. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part I. LNCS, vol. 4791, pp. 659–666. Springer, Heidelberg (2007)
16. Tenenbaum, J., de Silva, V., Langford, J.: A global geometric framework for nonlinear dimensionality reduction. Science 290 (2000)
17. Wein, W., Roper, B., Navab, N.: Integrating diagnostic B-mode ultrasonography into CT-based radiation treatment planning. IEEE Trans. on Med. Imaging 26(6), 866–879 (2007)
18. Wiskott, L., Sejnowski, T.: Slow feature analysis: Unsupervised learning of invariances. Neural Computation 14(4), 715–770 (2002)

Effects of Respiratory Liver Motion on Heating for Gated and Model-Based Motion-Compensated High-Intensity Focused Ultrasound Ablation

Erik-Jan Rijkhorst[1], Ian Rivens[2], Gail ter Haar[2], David Hawkes[1], and Dean Barratt[1]

[1] Centre for Medical Image Computing, University College London, London, UK
[2] The Institute of Cancer Research, Sutton, Surrey, UK

Abstract. Purpose: To quantify the effects of respiratory motion on high-intensity focused ultrasound heating of liver tissue by comparing the simulated ablation using a conventional respiratory gating versus a MR-model-based motion compensation approach.

Methods: To measure liver motion, dynamic free-breathing abdominal MR scans were acquired for five volunteers. Deformable registration was used to calculate continuous motion models, and tissue heating at a moving single focus was computed in 3-D by solving the bioheat equation. Ablated volume ratios with respect to the static case, $\mathcal{V}^{\mathrm{ab}}$, were determined for a range of exposure times t^{exp} and heating rates r.

Results: To achieve $\mathcal{V}^{\mathrm{ab}} > 90\%$ required $t^{\mathrm{exp}} < 0.5$s and $r > 120^{\circ}$C/s when gating, whereas $t^{\mathrm{exp}} < 1$s and $r > 60^{\circ}$C/s for motion-compensation.

Conclusions: Accurate compensation for respiratory motion is important for efficient tissue ablation. Model-based motion compensation allows substantially lower heating rates than gating, reducing the risk of skin burns and focal boiling.

Keywords: motion compensation, HIFU, heating, liver.

1 Introduction

High-intensity focused ultrasound (HIFU) ablation is being increasingly studied as a non-invasive treatment option for liver cancer [1–3]. However, liver respiratory motion can be up to a few centimetres under normal free-breathing conditions [4, 5]. Therefore, conventional gating methods lead to very lengthy treatment times, whereas to facilitate continuous ablation, some form of image guidance to steer the HIFU focus needs to be applied.

Recently reported MR-guided approaches show promising results [3], but these are logistically difficult and relatively expensive techniques, requiring specialist equipment. Ultrasound guidance based on direct speckle tracking may have limited application in practice due to the presence of ribs in the beam path [1, 2]. Moreover, the target location still needs to be identified during the procedure, typically using information from a pre-procedural CT or MR scan.

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 605–612, 2011.

The model-based motion compensation approach proposed in this paper overcomes some of these limitations by predicting liver motion using a continuous motion model. In this method, alignment of the patient's rigid anatomy with the model's co-ordinate system results in a known target location. In addition, temporal synchronisation of the motion model with the subject's breathing by means of a surrogate respiratory signal is required. This can be achieved, for instance, using a separate diagnostic ultrasound probe to track diaphragm motion during treatment at an oblique angle from below the ribs.

Although the need for real-time organ motion compensation is widely acknowledged, the impact of residual liver motion for different image-guidance schemes in terms of thermal dose has received little attention to date. The aim of this work was to compare the performance of a conventional gated approach with a model-based motion compensation method. To achieve the same ablated volume, gating requires shorter exposure times and higher heating rates than motion-compensated HIFU, but undesired effects such as skin burns and focal boiling may then become a significant limitation. To compute the required heating rates, continuous motion models using deformable image registration of dynamic MR sequences obtained from five human volunteers were calculated. Both gated and model-guided HIFU treatments in the presence of respiratory liver motion were simulated by solving the bioheat equation, and the thermal dose and corresponding ablated volumes were computed for a range of heating rates and exposure times.

2 Materials and Methods

2.1 Free-Breathing MR Scans

Free-breathing MR scans with a field-of-view covering the whole abdomen were acquired for five volunteers. A balanced-SSFP sequence (TR/TE=4.3/1.46 ms, 30° flip angle) was used to obtain high-resolution 4D dynamic scans during free breathing using a 1.5 T MR scanner (Philips Achieva, Best, The Netherlands) at Guy's Hospital, London. Parallel imaging with a 32-channel coil array using a SENSE acceleration factor of 4.6 resulted in scan times of approximately one second per dynamic volume. Since the largest liver motion normally occurs in the sagittal plane [4, 5], the highest reconstruction resolution of 1.4×1.4 mm was chosen in this plane, resulting in a slice thickness of 4 mm. For each volunteer, first a set of normal breathing scans were acquired, followed by a set obtained during deeper breathing. Each set contained a total of $N^{\mathrm{acq}} = 25$ volumes, covering ~ 4 breathing cycles and took ~ 20 s to acquire.

2.2 Deformable Registration of the Dynamic Scans

Tissue displacements throughout the liver were computed by registering each volume within a set of dynamic MR scans to a reference volume using a non-rigid fluid registration method [6]. This method solves the time-dependent Navier-Lamé equations for a compressible viscous fluid resulting in a diffeomorphic

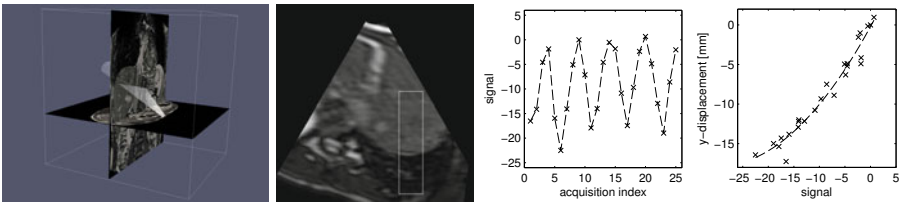


Fig. 1. Geometry of the virtual ultrasound sector used to re-slice the dynamic MR scans (1st panel). Example of a diaphragm navigator window manually located in a slice taken through the reference dynamic MR volume at the interventional ultrasound location (2nd panel) and corresponding surrogate respiratory signal (3rd panel). The 4th panel shows an example of a motion model fit of y-displacement data as a function of signal value for a single voxel location.

transformation. The registration is driven by image-derived forces and employs a full multi-grid scheme. The volume corresponding to the breathing phase closest to the time of full exhale was chosen as the reference volume. Registering this volume to all other volumes resulted in a set of N^{acq} displacement fields, denoted by $\boldsymbol{u}_i(\boldsymbol{r})$, with $\boldsymbol{r}$ denoting the 3-D voxel location within the volumes.

To assess the deformable registration accuracy, corresponding anatomical landmarks, such as vessels and vessel bifurcations, were picked manually in each dynamic scan, resulting in a set of N^{acq} points for each location. Landmark points were identified for each volunteer at five different locations distributed throughout the liver. The slice thickness in the lateral direction was 4 mm, and, because of the poor image quality of the dynamic scans, we did not attempt to quantify motion in the lateral direction. Consequently, only those landmarks clearly visible in a single sagittal plane over the complete dynamic sequence were chosen for the analysis. The target registration error (TRE) was calculated as the root-mean-square distance between the landmark points in the reference volume and all other volumes, following registration.

2.3 Surrogate Respiratory Signal

To synchronise a motion model with a subject's breathing during treatment, we assume a diagnostic ultrasound probe is used to acquire dynamic B-mode images of the moving diaphragm at an oblique angle from below the ribs. The motion model is computed using a corresponding diaphragm signal, obtained by re-slicing the dynamic MR volumes at the same location as used during treatment (see Fig. 1). A navigator window was placed over the dome of the right hemi-diaphragm, and successive navigator images were aligned to a reference navigator image at the time of full exhale. This was done using a 1-D translational registration method employing a gradient descent optimisation, and the mean of squared intensity differences as the similarity measure. The resulting 1-D displacements, s_i $(i = 1, \ldots, N^{\mathrm{acq}})$, were used as the surrogate respiratory signal (Fig. 1).

2.4 Liver Motion Models

To express continuous liver motion as a function of the respiratory signal s, a smooth function (see Fig. 1, rightmost panel) is fitted to the discrete registration derived displacements $\boldsymbol{u}$, denoted by

$$\boldsymbol{u}_i(\boldsymbol{r}) \equiv \boldsymbol{u}(\boldsymbol{r}, s_i) \ , \quad \text{with} \ \ i = 1, \ldots, N^{\text{acq}} \ . \tag{1}$$

Second order polynomials, given by

$$\hat{\boldsymbol{u}}(\boldsymbol{r}, s) = \boldsymbol{\alpha}_1(\boldsymbol{r}) + \boldsymbol{\alpha}_2(\boldsymbol{r})s + \boldsymbol{\alpha}_3(\boldsymbol{r})s^2 \ , \tag{2}$$

were used as the motion model fitting function to approximate the data sets. Note that the motion model coefficients $\boldsymbol{\alpha}$ are vector quantities defined at each voxel location $\boldsymbol{r}$. Inserting Eq. (1) into (2) results in an over-determined system of linear equations, given by

$$\begin{bmatrix} \boldsymbol{u}_1(\boldsymbol{r}) \\ \vdots \\ \boldsymbol{u}_{N^{\text{acq}}}(\boldsymbol{r}) \end{bmatrix} = \begin{bmatrix} 1 & s_1 & s_1^2 \\ \vdots & \vdots & \vdots \\ 1 & s_{N^{\text{acq}}} & s_{N^{\text{acq}}}^2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_1(\boldsymbol{r}) \\ \boldsymbol{\alpha}_2(\boldsymbol{r}) \\ \boldsymbol{\alpha}_3(\boldsymbol{r}) \end{bmatrix} \ . \tag{3}$$

From this, the motion model coefficients were determined by computing the Moore-Penrose pseudo-inverse using singular value decomposition.

2.5 Thermal Model of Tissue Heating

The evolution of the tissue temperature T as a function of time was computed in 3-D using the implicit Crank-Nicolson finite difference method for solving the bioheat equation [7, 8]

$$\rho C \frac{\partial T}{\partial t} = K \nabla^2 T - W_b C_b (T - T_b) + q(t) \ , \tag{4}$$

with ρ the mass density, C the specific heat, and K the thermal conductivity of the tissue. Cooling due to perfusion is modelled by the second term on the right hand side, with C_b the specific heat, W_b the perfusion rate, and T_b the temperature of the blood. Table 1 presents the material parameters which were kept constant in the simulations. The HIFU energy density $q(t)$ was computed from the intensity profile obtained using the Rayleigh integral for a spherical

Table 1. Material parameters used in the heating simulations [8, 10]

parameter description	value	unit
liver specific heat	$C = 3770$	J kg^{-1} °C^{-1}
liver thermal conductivity	$K = 0.508$	W m^{-1} °C^{-1}
blood perfusion rate	$W_b = 19.5$	kg m^{-3} s^{-1}
blood specific heat	$C_b = 3770$	J kg^{-1} °C^{-1}
blood temperature	$T_b = 37$	°C

bowl transducer [8] with a focal length of 15 cm and a radius of 5 cm. The dependence of the energy density q on time t in Eq. 4 denotes the effects of motion on heating. This was implemented by using the motion model coefficients $\boldsymbol{\alpha}$ from Eq. 3 to compute local displacements with which to translate q relative to the computational domain. To quantify heating effects we use the concept of thermal dose [9], given by

$$\mathrm{TD} = \sum_{t_0}^{t_{\mathrm{exp}}} R^{43-T} \Delta t \ , \tag{5}$$

with $R = 0.25$ for $T < 43°\mathrm{C}$, and $R = 0.5$ for $T > 43°\mathrm{C}$, and Δt the time step used in the heating simulations. Tissue was assumed to be fully ablated when $\mathrm{TD} \geq 240$ cumulative equivalent minutes.

2.6 Simulated Treatment Scenarios

Two scenarios for HIFU treatment during free-breathing were simulated: respiratory gating and continuous MR-model-based motion compensation. For both methods, respiratory liver motion as it occurs during an intervention was simulated using a subject-specific motion model, derived from a set of dynamic MR scans obtained during normal breathing. For the gating approach, ablation was gated symmetrically around the time of full exhale in the breathing cycle.

For the motion compensation method, the HIFU focus was assumed to be steered according to a model-based prediction of local liver motion. Although it would be more realistic to simulate focus steering using a transducer array, for this initial study we opted to simply translate the heating pattern generated by a single bowl transducer. The prediction model was computed from a second set of dynamic MR scans, obtained during deeper breathing a few minutes after the first set was acquired. In practice, the prediction model can be synchronised during treatment to the subject's breathing by means of a navigator signal obtained using real-time ultrasound imaging of the diaphragm (see Fig. 2). In our simulations motion correction was implemented by subtracting the predicted motion from the simulated subject motion, resulting in residual guidance errors representative of those that may occur in practice.

To take into account variations in motion occurring in different parts of the liver, three different simulated ablation locations were chosen. The treatment depth was set to 5 cm in the anterior-posterior direction, and the location of

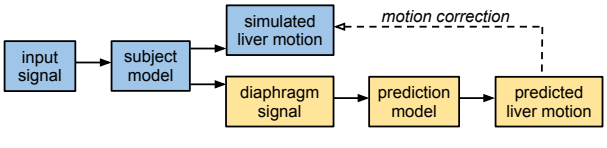


Fig. 2. Schematic of computations performed for the MR-model-based motion compensation scenario

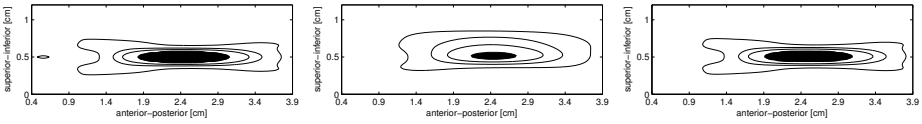


Fig. 3. Temperature contour plots of the temperature at $t^{\exp} = 1.5$ s of a single focus ablation with a heating rate of $r = 40^{\circ}\mathrm{C\,s}^{-1}$. The outer contour is at 40°C, with a 5°C spacing between contours. The filled contour shows the extent of the lesion corresponding to a thermal dose TD ≥ 240 minutes. Shown are the static case (left), gated (middle), and motion-compensated approach (right).

the focus was centred on the extent of the liver in the lateral direction. The locations were then set to 20%, 50%, and 80% of the extent of the liver in the inferior-superior direction. For each subject, simulations of single focus HIFU heating were performed for four different combinations of heating rate $r = q/\rho C$ and exposure time $t^{\exp}$, for both the gated and motion-compensated approaches. In all simulations, a sinusoidal function with a breathing period of 4 seconds, representative of actual observed subject breathing, was used as an input signal for the subject model. The ablated volume ratio with respect to the static case, denoted by $\mathcal{V}^{\mathrm{ab}}$, was determined for all three ablation locations within the liver.

3 Results

By analysing the registration results at the anatomical landmark locations for all volunteers, we found a mean displacement of 7.4 mm with a standard deviation of 5.9 mm, and a maximum displacement of 27.2 mm. After registration, the mean TRE was 1.7 mm, with a standard deviation of 1.2 mm, and a maximum of 8.4 mm. These results indicate that the registration accurately captures the respiratory motion from the sets of dynamic MR scans. As an example, Fig. 3 shows heating contour plots for a single focus ablation after 1.5 seconds of heating at a rate of $40^{\circ}\mathrm{C\,s}^{-1}$. For this particular case, the gated approach is insufficient to produce a properly ablated region, as can be seen from the spread of the temperature profile and much smaller ablated volume. In contrast, the motion-compensated method almost fully recovers the ablated volume of the static case. Table 2 presents the motion amplitudes at the three chosen treatment locations for all subjects as computed from the normal breathing motion models, together with the residual amplitudes after this motion was corrected for using the deeper breathing guidance motion models. Table 3 summarises the results for the ablated volume ratios $\mathcal{V}^{\mathrm{ab}}$ relative to the equivalent static cases for the gated and motion-compensated scenarios for four combinations of heating rate and exposure time. For each volunteer, results are given for three different locations, distributed over the superior-inferior extent of the liver. To achieve an ablated volume ratio $\mathcal{V}^{\mathrm{ab}} > 90\%$ relative to the static case, required an exposure time $t^{\exp} < 0.5$ s and a heating rate $r > 120^{\circ}\mathrm{C/s}$ for the gated approach, whereas for the motion-compensated method $t^{\exp} < 1$ s and $r > 60^{\circ}\mathrm{C/s}$.

Table 2. Maximum motion amplitudes in mm occurring during the simulations with the longest exposure times of $t^{\exp} = 2$ s, for all subjects. Amplitudes of simulated respiratory motion as they would occur during gated treatments are given in the first row. The second row shows residual displacement amplitudes after a second motion model was used to compensate for the subject motion.

	subject 1			subject 2			subject 3			subject 4			subject 5		
point location	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
gated approach	9.1	6.1	6.5	7.5	5.6	4.9	8.0	7.1	6.5	6.5	5.8	6.4	9.2	6.7	6.5
motion-compensated	1.5	0.1	0.7	2.1	1.8	2.2	1.0	0.9	0.5	3.3	3.2	2.7	0.6	0.8	1.0

Table 3. Ablated volume ratios $\mathcal{V}^{\text{ab}}$ expressed as a percentage relative to the equivalent static case for each subject, are given as a function of heating rates r and total exposure times $t^{\exp}$ for three different locations within the liver. Results for conventional gated and MR-model-based motion compensation approaches are shown.

	heating rate	exposure time	subject 1 locations			subject 2 locations			subject 3 locations			subject 4 locations			subject 5 locations		
	r (°C/s)	$t^{\exp}$ (s)	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
conventional gated approach	30	2.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	40	1.5	0	18	11	2	23	34	0	8	13	16	25	16	0	16	16
	60	1.0	50	76	71	64	80	85	62	71	74	77	81	77	52	75	75
	120	0.5	95	97	96	96	97	97	94	97	97	98	97	97	95	98	98
model-based motion compensation	30	2.0	64	99	90	46	47	31	92	84	91	0	13	17	91	87	80
	40	1.5	88	100	91	81	82	76	97	95	99	51	67	70	96	95	94
	60	1.0	94	99	97	95	95	94	95	97	98	90	92	93	98	98	96
	120	0.5	97	100	98	98	97	97	96	96	97	98	97	97	99	98	98

4 Discussion and Conclusions

We have presented detailed simulations of HIFU heating in the presence of realistic respiratory liver motion, for both gated and MR-model-based motion-compensated ablation. For the gated simulations, we found that the ablated volume increases when shorter HIFU exposure times, and correspondingly higher heating rates, were used. This is intuitive, since the target moves less when shortening the exposure time, and therefore less spread occurs in the applied heating pattern. Furthermore, from the results of Tables 2 and 3, the largest liver motion occurred at the most superior location in the liver, resulting in consistently smaller ablated volumes than for more inferior located points in the liver. For the motion-compensated method, the ablated volume sizes are much less sensitive to variations in exposure time and heating rate, as long as $t^{\exp} < 1$ s and $r > 60$°C/s.

Although the use of a simple spherical bowl transducer is sufficient for the purpose of this initial study, we are planning to incorporate more realistic phased array pressure fields to simulate beam steering while compensating for the effects

of ribs in the beam path. We conclude that model-based motion compensation allows for two times lower heating rates than gating, thereby reducing the risk of negative side-effects such as skin burns and focal boiling.

Acknowledgements. We would like to thank Tobias Schäffter for performing the MR scanning, and all volunteers for participating in this study. This work was supported by the EPSRC, grant numbers EP/F025750/1 and EP/F029217/1.

References

1. Pernot, M., Tanter, M., Fink, M.: 3-D real-time motion correction in high-intensity focused ultrasound therapy. Ultrasound in Medicine & Biology 30(9), 1239 (2004)
2. de Oliveira, P.L., de Senneville, B.D., Dragonu, I., Moonen, C.T.W.: Rapid motion correction in MR-guided high-intensity focused ultrasound heating using real-time ultrasound echo information. NMR in Biomedicine 23(9), 1103–1108 (2010)
3. Ries, M., de Senneville, B.D.D., Roujol, S., Berber, Y., Quesson, B., Moonen, C.: Real-time 3D target tracking in MRI guided focused ultrasound ablations in moving tissues. Magnetic Resonance in Medicine (2010)
4. von Siebenthal, M., Székely, G., Gamper, U., Boesiger, P., Lomax, A., Cattin, P.: 4D MR imaging of respiratory organ motion and its variability. Physics in Medicine and Biology 52(6), 1547–1564 (2007)
5. Rijkhorst, E.-J., Heanes, D., Odille, F., Hawkes, D., Barratt, D.: Simulating dynamic ultrasound using MR-derived motion models to assess respiratory synchronisation for image-guided liver interventions. In: Navab, N., Jannin, P. (eds.) IPCAI 2010. LNCS, vol. 6135, pp. 113–123. Springer, Heidelberg (2010)
6. Crum, W.R., Tanner, C., Hawkes, D.J.: Anisotropic multi-scale fluid registration: evaluation in magnetic resonance breast imaging. Physics in Medicine and Biology 50(21), 5153–5174 (2005)
7. Pennes, H.H.: Analysis of tissue and arterial blood temperatures in the resting human forearm. Journal of Applied Physiology 1(2), 93–122 (1948)
8. Meaney, P.M., Clarke, R.L., ter Haar, G.R., Rivens, I.H.: A 3-D finite-element model for computation of temperature profiles and regions of thermal damage during focused ultrasound surgery exposures. Ultrasound in Medicine & Biology 24(9), 1489–1499 (1998)
9. Sapareto, S.A., Dewey, W.C.: Thermal dose determination in cancer therapy. International Journal of Radiation Oncology*Biology*Physics 10(6), 787–800 (1984)
10. Williams, L.R., Leggett, R.W.: Reference values for resting blood flow to organs of man. Clin. Phys. Physiol. Meas. 10, 187 (1989)

Multiple Structure Tracing in 3D Electron Micrographs*

Vignesh Jagadeesh, Nhat Vu, and B.S. Manjunath

Department of ECE and Center for Bioimage Informatics
University of California, Santa Barbara CA - 93106

Abstract. Automatic interpretation of Transmission Electron Micrograph (TEM) volumes is central to advancing current understanding of neural circuitry. In the context of TEM image analysis, tracing 3D neuronal structures is a significant problem. This work proposes a new model using the conditional random field (CRF) framework with higher order potentials for tracing multiple neuronal structures in 3D. The model consists of two key features. First, the higher order CRF cost is designed to enforce label smoothness in 3D and capture rich textures inherent in the data. Second, a technique based on semi-supervised edge learning is used to propagate high confidence structural edges during the tracing process. In contrast to predominantly edge based methods in the TEM tracing literature, this work simultaneously combines regional texture and learnt edge features into a single framework. Experimental results show that the proposed method outperforms more traditional models in tracing neuronal structures from TEM stacks.

Keywords: Tracing, CRFs, Textures.

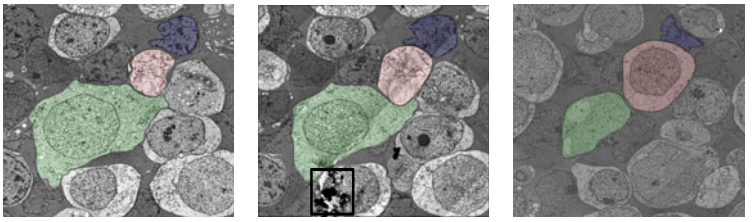
1 Introduction

Understanding the interconnectivity structure of the brain is a grand challenge in neuroscience. Recent developments in imaging have enabled capturing massive amounts (in terabytes) of Transmission Electron Micrograph (TEM) data at sub-nanometer resolutions. Manual analysis of these data repositories is infeasible, justifying the need for evolving image analysis algorithms. One of the challenges in interpreting these data repositories lies in automated tracing of multiple interacting neuronal structures in 3D. This paper proposes a robust and efficient tracing algorithm employing conditional random fields(CRFs).

Problem Definition: Images displayed in Figure 1 correspond to different z-slices from a TEM stack. Three sample structures are shaded to illustrate difficulty levels in tracing: simple(red), medium(green) and hard(blue). As can be seen, the structures express regional textures that are discriminative and noisy.

* This work was supported by NSF OIA 0941717. The authors thank Dr.Robert Marc, Dr.Brain Jones and Dr.James Anderson from the Univ. of Utah for providing data used in experiments and for useful discussions.

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 613–620, 2011.



(a) (b) (c)

Fig. 1. (Best Viewed in Color) Three representative slices with manual annotations. Structure in red is simple to trace since it has well defined gradients and regional cues. Structure in green is tougher since it shares boundaries with many different neighbors and a variety of gradient profiles arise on its border. Structure in blue is toughest to trace due to inconsistent region and edge information.

Furthermore, each structure has special gradient profiles that cannot be captured using simple gradient operators. Finally, structures could deform considerably without arbitrary movement from one slice to another. The figure also illustrates problems introduced during the imaging process (sample damage shown by a black box in Figure 1(b), illumination artifacts from Figure 1(b) to 1(c)). As is evident from Figure 1, it would be desirable to model regional cues (intensity, texture etc ..), characteristic gradient profiles (learnt edges) and 3D smoothness constraints (dynamics of structures in the z-direction) in a unified framework that implicitly handles arbitrary changes in topology. This work aims to convince the reader that CRFs are capable of exploiting multiple information sources for tracing neuronal structures in TEM stacks.

Contributions: The primary contribution of this work is in designing an algorithm for tracing multiple interacting 3D structures in TEM stacks. Salient features of the proposed model include:

- Semi-supervised edge learning scheme for propagating high confidence edge maps during the tracing process (Section 2.1)
- Design of a higher order CRF cost for simultaneously enforcing 3D label smoothness and capturing regional textures (Section 2.2)

Related Work: While there has been a lot of interest in EM tracing over the past year, this work differs from existing literature [3],[6],[5],[4] (and references therein) in the following ways. The nature of data considered is vastly different, in that it has discriminative regional texture in addition to characteristic gradient profiles. Most existing techniques approach the problem employing strong edge learning methods since the datasets considered present edge detection as the important challenge. In particular [7] propose a global technique that performs optimization over the entire stack. However, scaling global techniques to large datasets would be difficult. Sample damage and unfavorable imaging conditions could further seriously affect a truly global approach. This motivates the proposed approach where multiple interacting structures are traced in a scalable manner.

The proposed solution adopts a CRF framework for integrating multiple sources of information described previously. The following paragraph provides a brief background and introduces notations that will be used in this paper. CRFs are image models that capture contextual interaction between pixels. Given features x_{p^z} of a pixel p in slice z of the stack, the goal is to infer its label $y_{p^z} \in \mathcal{L} = \{1, 2..L\}$. $\mathcal{L}$ is the label set containing L labels, which in the present case is the number of structures to be traced. $\mathcal{P}^z$ denotes the set of pixels constituting slice z. Labeling is achieved by minimizing a CRF energy comprising unary, interaction and higher order terms, see Equation 1. Unary potentials encode the likelihood of a pixel p^z to take on label y_{p^z}. Interaction potentials encourage label smoothness between p^z and its neighbors q^z contained in a neighborhood system $\mathcal{N}_{p^z}$. Higher order terms encourage label consistency across any clique (group of pixels) c^z contained in the set of cliques $\mathcal{C}^z$ in slice z. For this work, cliques are superpixels generated by oversegmentation. The CRF energy is given by:

$$E(y_{p^z}) = \underbrace{\sum_{p^z \in \mathcal{P}^z} V_p(y_{p^z})}_{Unary\ Potential} + \underbrace{\sum_{p^z \in \mathcal{P}^z, q^z \in \mathcal{N}_{p^z}} V_{pq}(y_{p^z}, y_{q^z})}_{Interaction\ Potential} + \underbrace{\sum_{c^z \in \mathcal{C}^z} V_c(y_{c^z})}_{Higher\ Order\ Term} \quad (1)$$

The above energy function can be efficiently minimized using graph cuts if each term obeys a submodularity constraint. The time complexity of inference employed in this work is similar to traditional alpha expansions. For further details [9], [2], [1] are comprehensive sources of reference.

This paper is organized as follows. The second section describes the construction of potential functions for the proposed higher order CRF model. The subsequent section presents experimental results on TEM stacks with a quantitative analysis. The final section concludes the paper with a discussion on future work.

2 Multiple Structure Tracing with Higher Order CRFs

The proposed model is obtained by constructing unary, interaction and higher order terms in Equation 1. The unary potentials are coarse object/background likelihoods computed using information from previous segmentations. The unary potentials are unregularized and require interaction models (second and higher order) for smooth segmentation. Traditional second order interaction models comprise first order gradient operators, which are often incapable of capturing a wide range of edge profiles. In contrast, the proposed semi-supervised scheme for edge propagation captures a wider range of edge profiles using a learnt model as shown in Figure 2. The edge propagation scheme does not assume smoothness across the z-direction in its construction. However, there are scenarios where unary and pairwise terms become unreliable (for the hard structure in Figure 1(c) and in case of sample damage). In such scenarios, a model for capturing higher order regional interactions for resisting failure caused due to first and second order models, and for enforcing smoothness across the third dimension is required. The robust Pn model is employed for this purpose.

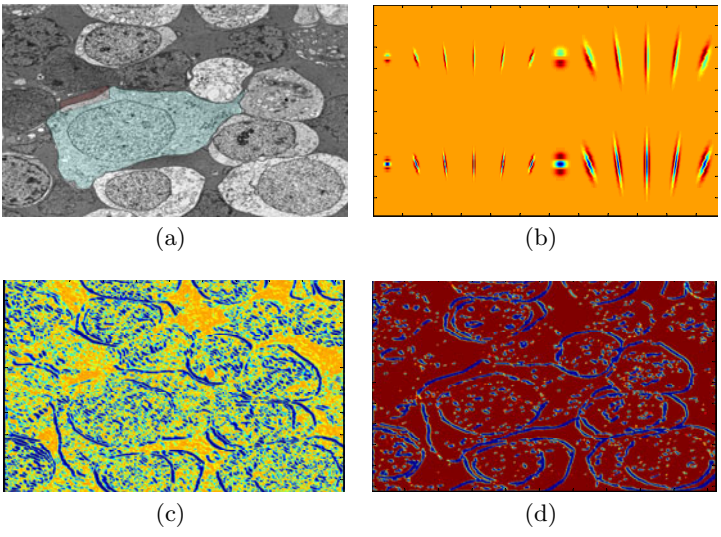


Fig. 2. (Best viewed in color, red pixels have high values/blue pixels have lower values) (a) Green region is to be traced. Red region in (a) shows a sample edge that needs to be detected. (b) Filterbank used (c) Traditional second order energy, gradients not well localized and edges in the red region shown in (a) are missed. (d) Proposed scheme provides accuracy localization for structure of interest. Observe red region in (a) is detected with high confidence.

2.1 Semi-supervised Edge Propagation

Interaction potentials are given by $V_{pq}(y_{p^z}, y_{q^z}) = \lambda_I \exp(-\frac{(I_{p^z} - I_{q^z})^2}{2\sigma_I^2}) \frac{1}{dist(p^z, q^z)}$ $\delta(y_{p^z} = y_{q^z})$. λ_I, σ_I are parameters controlling influence of interaction potentials and edge quality respectively, δ is a dirac delta function evaluating to 0 if the condition in parenthesis is satisfied and 1 otherwise. Computing V_{pq} in the above fashion is equivalent to a first order gradient operator, which is not suitable for TEM datasets. This work leverages contour initialization as partial labeling and propagates edges through the stack using a semi supervised scheme. The idea is to learn edge textons [10] to improve structure specific edge detection. This is in contrast to recent efforts that learn application specific edges using large amounts of training data. For ease of explanation, a two label problem with the goal of propagating edges between slices $z-1$ and z is considered.

The filter bank considered comprises of M filters and is given by $\mathcal{F} = [F_1 F_2 ... F_M]$, see 2(b). Specifically, the filter bank has twelve first and twelve second derivative filters, each at six orientations and two scales (a subset of the Leung Malik filter bank). The labels $y_{p^{z-1}}$ are known and so are edge pixels for slice $z-1$. Initially the images are convolved with $\mathcal{F}$ resulting in filter responses $R^i_{p^{z-1}}$ and $R^i_{p^z}$ for slices $z-1$ and z respectively. The intuition behind using a filter bank is capture varied gradient profiles exhibited by the individual filters, in contrast to simple first order gradients. Since there are M filter responses, concatenation

of all responses $R^i_{p^{z-1}}, \forall i \in [1,2,..M]$ at a pixel p^{z-1} is represented by the M dimensional vector $\mathbf{R}_{p^{z-1}}$. Subsequently, the edge filter responses are clustered in an M dimensional feature space for the the K edge textons $\mathbf{C}_k$. The texton indices t_{p^z} are found by performing a nearest neighbor search of filter responses on the image in slice z with the textons $\mathbf{C}_k$. The feature descriptor for training classifiers are local histograms of texton indices, $h_{p^z}(k) = \sum_{q^z \in \mathcal{N}_{p^z}} \delta(t_{q^z} \neq k), \forall k \in [1,2,...K], h_{p^z} \in \mathcal{R}^K$. A naive Bayes classifier is trained using the feature vectors $h_{p^{z-1}}$ and tested on features from the current frame h_{p^z}. The output of the classifier is a posterior probability $Pr(e_{p^z})$ that yields large values at locations where there is a high probability of the structure specific edge. The interaction potentials for the conditional random field can now be rewritten as $V_{pq}(y_{p^z}, y_{q^z}|y^{z-1}) = \lambda_I(1 - \frac{Pr(e_{p^z})+Pr(e_{q^z})}{2})\delta(y_{p^z} = y_{q^z})$.

2.2 Robust Higher Order Potentials

It is well known that higher order CRFs are capable of modeling larger spatial interactions. These models are ideal for capturing textures inherent in the neuronal structures. This work adopts the recently proposed Robust P^n model[8], a class of potential functions that are a strict generalization of the Potts model for second order interactions. The idea behind the model is that pixels constituting a superpixel (homogenous regions) are more likely to take the same label. The cost is expressed as $V_c(y_{c^z}) = \min\{\min_{k \in \mathcal{L}}((|c^z| - n_k(y_{c^z}))\theta_k + \gamma_k), \gamma_{max}\}$, where $|c^z|$ is the cardinality of the clique in slice z, $\mathcal{L}$ is the label set containing the set of possible labels, $n_k(y_{c^z})$ is the number of labels in the clique taking label $k \in \mathcal{L}$, $\theta_k = \frac{\gamma_{max}-\gamma_k}{Q}$ and Q is a truncation parameter controlling magnitude of label violations in the clique. γ_k, γ_{max} are penalties associated with the clique taking label k and mixed labeling respectively. γ_k is usually set to zero since uniform superpixel labelings are not penalized, $\gamma_{max} = |c|^{\theta_\alpha}(\theta^h_p + \theta^h_v G(c))$, where $\theta_\alpha, \theta^h_p, \theta^h_v$ are free parameters and $G(c^z)$ is a quantity indicative of superpixel quality. While enforcing label consistency, it is imperative that smoothness across the z-direction is preserved. One requires a cost $V_c(y_{c^z}|y^{z-1})$ and a method for evaluating G_c for superpixels.

Intuition: If a contour with label l propagates down till slice $z-1$, then the superpixels in slice z overlapping with the contour in slice $z-1$ are most likely to take the label l. In other words, the dominant label for superpixel c^z (overlapping with contour labelled l in slice $z-1$)is l. If all pixels constituting the superpixel take on the dominant label, minimal penalty is incurred. On the other hand, as the number of pixels violating the dominant label increases, higher cost (closer to γ_{max}) is incurred.

Modified Higher Order Cost: The segmentation of the current slice needs to respect the label homogeneity of the current slice, and also preserve smoothness across the z-direction. The proposed model incorporates smoothness across the z-direction using the variables γ_k. In previous constructions, the value of γ_k was usually set to zero since label homogeneity indicated smooth segments in 2D.

However, label homogeneity across 2D does not directly imply smoothness across 3D. In particular, an unsymmetrical Dice coefficient is used for measuring the overlap of a certain superpixel (c^z) to its overlapping contour in slice $z-1$.

$$n_{c^z}^k = \frac{\sum_{p^z \in P^z} M_{p^z}^{c^z} \wedge \delta(\hat{y}_{p^z} \neq k)}{\sum_{p^z \in P^z} M_{p^z}^{c^z}}, \quad M_{p^z}^{c^z} = \begin{cases} 1, & \forall p^z \in c^z, \\ 0, & otherwise \end{cases}$$

In the above equation $\hat{y}_{p^z} = y_{p^{z-1}}$, meaning that label predictions for slice z are labels propagated from slice $z-1$. Smoothness across the third dimension can now be incorporated as $\gamma_k \propto 1 - \frac{n_{c^z}^k}{|c^z|}, \quad c^z \in \mathcal{C}^z, \quad \forall k \in \mathcal{L}$. The constraint $\gamma_k < \gamma_{max}$ is always enforced for the cost to be minimized by graph cuts. Superpixel quality is evaluated using the variance of local intensity features on the superpixel. The modified higher order cost is obtained by substituting for γ_k, $V_c(y_{c^z}|y^{z-1}) = \min\{\min_{k \in \mathcal{L}}((|c^z| - n_k(y_{c^z}))\theta_k + \gamma_k), \gamma_{max}\}$.

2.3 First Order Potentials

The unary potential models the likelihood of a certain pixel taking up label $l \in \mathcal{L} = \{1, 2, ..L\}$. Electron Micrograph data is rich in texture, but not of the sort one would find in traditional texture analysis literature. It is used as a valuable cue by biologists, but seems to contain a lot of noise. A local multiscale feature similar to texture histograms, $I_z^i = I_z * g_{\sigma_i}, \quad 1 \leq i \leq N_f$ where $N_f = 3$ is employed. The above equation refers to smoothing of image I_z at position z on the stack by a Gaussian kernel g_{σ_i} with variances σ_i. Concatenation of filter responses at each pixel yields a feature vector in $\mathcal{R}^{N_f}$ and likelihoods are obtained by standard histogram backprojection techniques. The overall unary potential can be expressed as: $V_p(y_{p^z}|y_{p^{z-1}}, I_z) = -\log(Pr(I_z|y_{p^z}, y_{p^{z-1}})Pr(y_{p^{z-1}}|y_{p^z}))$. Note that $Pr(I_z|y_{p^z}, y_{p^{z-1}})$ is obtained by backprojecting the multiscale histograms, and $Pr(y_{p^{z-1}}|y_{p^z})$ is a signed distance function constraining the contour to be close to its position in the previous slice (shape prior). Further, optical flow fields could also be employed if there are mild registration errors between slices for simultaneous segmentation and registration.

3 Experimental Results

Experimental results are reported on Electron Micrographs of the retina. Neuronal structures were traced for over 45 slices of the stack. Results are compared with ground truth for a quantitative analysis of pixel errors.

Single Structure Tracing: Tracing single structures is performed to provide a proof of concept for the edge learning and higher order models. Tracing was done by providing an initial contour on the first frame of the stack, and is quantified by computing the F-measure (where P and R are precision and recall), $F = \frac{2PR}{P+R}$. As can be observed from Figure 3(b), traditional second order terms get distracted by noisy gradients and were not able to recover once the contour was

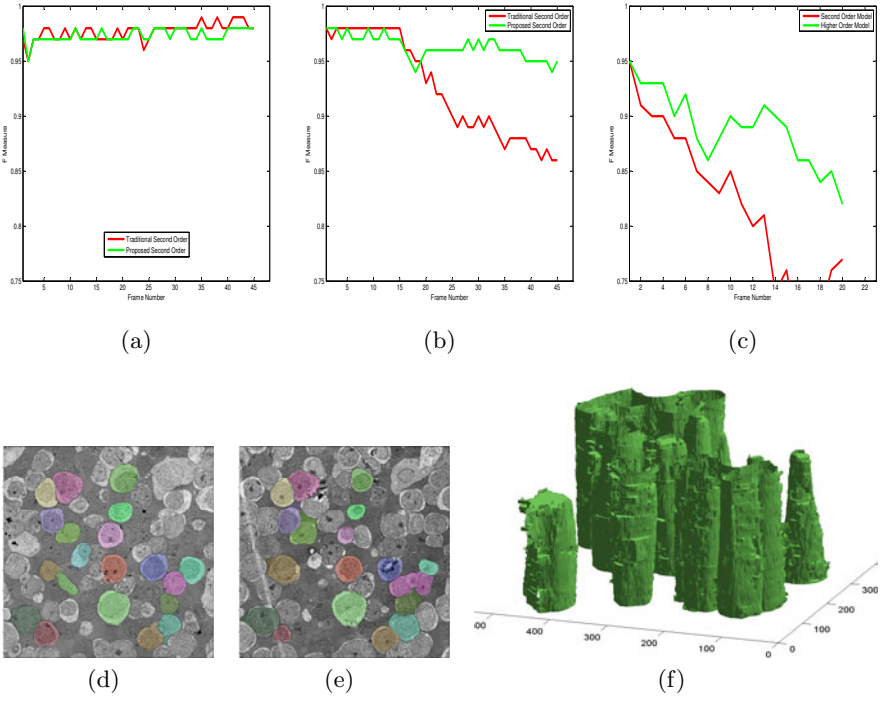


Fig. 3. (Best Viewed in Color) (a) Tracing on simple structure (shown in 1) (b) Performance on medium difficulty structure. (c) Performance on difficult target. (d) and (e) Sample tracing results on different z slices. (f) 3D reconstruction of 20 traced structures.

Table 1. F measures for multiple structure tracing

#Slices	#Contours	Traditional	Proposed
10	300	0.892	0.906
20	600	0.852	0.872
30	900	0.800	0.823

lost. The error on subsequent slices were additive, leading to poor performance as shown by the red lines. However, the second order model with proposed edge learning was able to resist distractions of noisy gradients as can be seen from green lines. On the medium difficulty structure, average F-measure for the traditional scheme was 0.925, in comparison to 0.962 for the proposed scheme. Figure 3(c) illustrates an example using the difficult target where the higher order model was able to outperform simple second order interactions. On the hard structure, average F-measure for the second order model was 0.817, while the higher order model yielded 0.889.

Multiple Structure Tracing: Experiments were also performed for multiple structure tracing in 3D. A total of 30 interacting structures were traced in parallel. The performance of the algorithm is promising (see Figure 3(f)) with quantitative results in Table 1. Losing the trace of a structure usually happens when there are arbitrary appearance variations (also caused due to illumination artifacts).

Conclusions: This paper presented a novel framework for tracing multiple neuronal structures in TEM stacks. Experimental results were presented on data from TEM stacks for single and multiple structure tracing. Future work includes investigation of techniques to scale up tracing to large datasets and testing stability of the algorithm under imperfect initialization.

References

1. Boykov, Y., Kolmogorov, V.: An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. PAMI 26(9), 1124–1137 (2004)
2. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. PAMI 23(11), 1222–1239 (2002)
3. Chklovskii, D., Vitaladevuni, S., Scheffer, L.: Semi-automated reconstruction of neural circuits using electron microscopy. Curr. Op. In: Neurobiology (2010)
4. Jain, V., Bollmann, B., Richardson, M., Berger, D., Helmstaedter, M., Briggman, K., Denk, W., Bowden, J.: Boundary learning by optimization with topological constraints. In: CVPR 2010, pp. 2488–2495. IEEE, Los Alamitos (2010)
5. Jurrus, E., Hardy, M., Tasdizen, T., Fletcher, P., Koshevoy, P., Chien, C., Denk, W., Whitaker, R.: Axon tracking in serial block-face scanning electron microscopy. Medical image analysis 13(1), 180–188 (2009)
6. Kaynig, V., Fuchs, T., Buhmann, J.: Neuron geometry extraction by perceptual grouping in sstem images. In: CVPR 2010, pp. 2902–2909. IEEE, Los Alamitos (2010)
7. Kaynig, V., Fuchs, T.J., Buhmann, J.M.: Geometrical consistent 3D tracing of neuronal processes in ssTEM data. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6362, pp. 209–216. Springer, Heidelberg (2010), http://dx.doi.org/10.1007/978-3-642-15745-5_26, doi:10.1007/978-3-642-15745-526
8. Kohli, P., Ladickỳ, L., Torr, P.: Robust higher order potentials for enforcing label consistency. IJCV 82(3), 302–324 (2009)
9. Kolmogorov, V., Zabih, R.: What energy functions can be minimizedvia graph cuts? PAMI, 147–159 (2004)
10. Leung, T., Malik, J.: Representing and recognizing the visual appearance of materials using three-dimensional textons. IJCV 43(1), 29–44 (2001)

Neural Process Reconstruction from Sparse User Scribbles

Mike Roberts[1], Won-Ki Jeong[1], Amelio Vázquez-Reina[1], Markus Unger[2], Horst Bischof[2], Jeff Lichtman[1], and Hanspeter Pfister[1]

[1] Harvard University
[2] Graz University of Technology

Abstract. We present a novel semi-automatic method for segmenting neural processes in large, highly anisotropic EM (electron microscopy) image stacks. Our method takes advantage of sparse scribble annotations provided by the user to guide a 3D variational segmentation model, thereby allowing our method to globally optimally enforce 3D geometric constraints on the segmentation. Moreover, we leverage a novel algorithm for propagating segmentation constraints through the image stack via optimal volumetric pathways, thereby allowing our method to compute highly accurate 3D segmentations from very sparse user input. We evaluate our method by reconstructing 16 neural processes in a $1024 \times 1024 \times 50$ nanometer-scale EM image stack of a mouse hippocampus. We demonstrate that, on average, our method is 68% more accurate than previous state-of-the-art semi-automatic methods.

1 Introduction

Mapping neural circuitry is an important ongoing challenge in neurobiology. Current approaches to this task involve tracing neural processes through segmented nanometer-scale EM (electron microscopy) image stacks of brain tissue. Since our understanding of neural circuitry is often limited by our ability to reconstruct neural processes from EM image stacks, accurately segmenting neural processes is an important open problem in the medical image analysis community.

Dense reconstruction algorithms [1, 7, 11–13, 15, 20] generally rely on supervised learning methods to automatically classify every pixel in an image stack according to the type of cellular structure to which it belongs. However, no dense reconstruction algorithm can reliably produce segmentations that are completely free of topological errors. In practice, these methods often require significant user effort to correct errors in the automatically generated segmentations.

On the other hand, *sparse reconstruction algorithms* rely on the user to interactively guide the segmentation of individual neural processes. Most existing sparse algorithms compute 3D reconstructions as sequences of locally optimal 2D segmentations after the user provides an initial 2D contour [4, 8, 10, 14, 19]. However, these approaches do not optimally enforce 3D geometric consistency constraints on the resulting segmentation, and therefore often require frequent user intervention. The recent *Markov Surfaces* algorithm [16] requires user-defined

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 621–628, 2011.

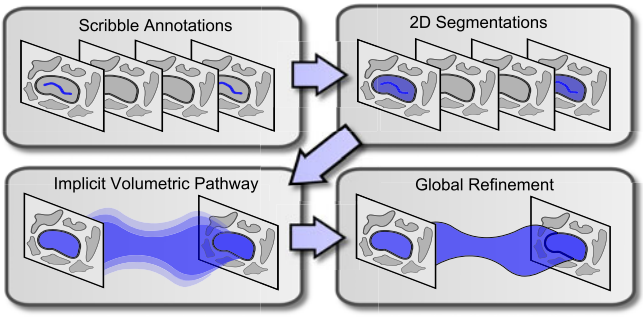


Fig. 1. Overview of our method. We assume that we are given scribble annotations indicating a neural process of interest on the first and last slices of an image stack (top left). We compute 2D segmentations that contain the scribble annotations and align with strong image edges; these 2D segmentations define hard constraints on our 3D segmentation (top right). We propagate the 2D segmentations through the image stack according to an implicitly represented volumetric pathway, which we compute based on the dense optical flow between image slices; the interior level sets of this volumetric pathway define soft constraints on our 3D segmentation (bottom left). We compute the final 3D segmentation by globally refining the volumetric pathway according to an anisotropic variational segmentation model that aligns with strong in-plane image edges and enforces 3D smoothness (bottom right).

2D contours on the first and last slices of an image stack. This algorithm automatically tessellates a set of globally optimal paths between these contours, relying on 2D Bézier interpolation to produce smooth surfaces. However, since Bézier interpolation does not take into account the underlying image data, the resulting segmentations may ignore important image features.

In this paper, we introduce a novel method for neural process reconstruction that only requires very sparse scribble annotations as input (Fig. 1). We evaluate our method by reconstructing 16 neural processes in a $1024 \times 1024 \times 50$ nanometer-scale EM image stack of a mouse hippocampus. We demonstrate that, on average, our method is 68% more accurate than Markov Surfaces [16], 91% more accurate than Geo-Cuts [2], and 263% more accurate than Marker-Controlled Watersheds [6].

2 Our Method

We observe that the problem of reconstructing neural processes through highly anisotropic EM image stacks is conceptually similar to the problem of tracking moving objects in video sequences. Based on this observation, our work is inspired by the recent *Anisotropic Total Variation* model proposed by Unger *et al.* [17], which tracks objects through video sequences based on sparse constraints provided by the user. However, the absence of color information in EM image data and poor spatial continuity across EM image slices prevent the direct application of this method to neural process reconstruction (Fig. 2a). We observe

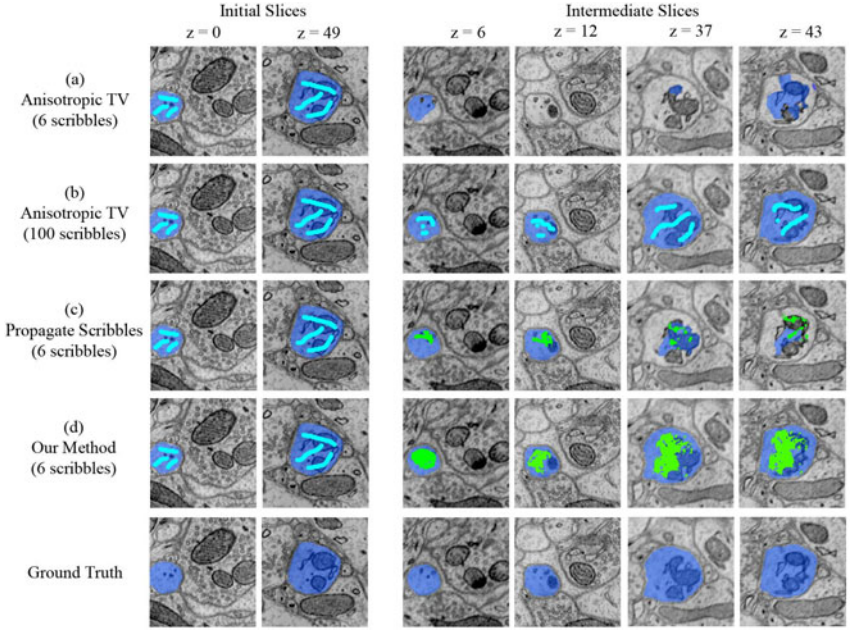


Fig. 2. Key observations motivating our method. Anisotropic Total Variation [17] fails to segment this neural process from sparse scribble annotations (a), but succeeds if scribble annotations are given on every slice (b). Our method only requires scribble annotations on the first and last slices because we automatically propagate segmentation constraints through the image stack. However, propagating scribble annotations as segmentation constraints results in a significant under-segmentation of this neural process (c). Instead, we compute 2D segmentations from the scribble annotations and propagate the 2D segmentations, resulting in an accurate segmentation of this neural process (d). Scribble annotations are shown in light blue, segmentations are shown in dark blue, and automatically propagated segmentation constraints are shown in green.

that this model can robustly track the neural process of interest if the user provides constraints on each slice of the image stack (Fig. 2b). This observation motivates our method for automatically propagating segmentation constraints through the image stack based on the dense optical flow between slices.

If we propagate the user-provided scribble annotations through the image stack as soft segmentation constraints, we observe an uneven distribution of propagated constraints. This can result in a significant under-segmentation of the neural process of interest (Fig. 2c). In contrast, if we propagate accurate 2D segmentations (instead of scribble annotations) through the image stack, we observe more evenly distributed segmentation constraints. This results in an accurate 3D segmentation of the neural process of interest (Fig. 2d). These observations motivate our method for computing 2D segmentations and subsequently propagating them through the image stack as soft segmentation constraints.

Input. We assume that the user marks the neural process of interest with a few foreground and background scribbles on the first and last image slices. For the sake of notational clarity, we assume that the neural process of interest is roughly orthogonal to the image stack. In practice, we allow the user to segment neural processes that run in any direction by projecting the segmentation problem along any 1D path through the image volume.

Computing 2D Segmentations. We compute 2D segmentations of the neural process of interest by globally minimizing the following variational segmentation energy [18]:

$$\underset{u_i}{\operatorname{argmin}} \int_{\mathbf{x} \in \Omega_i} (g_i |\nabla u_i| + f_i u_i)\, d\mathbf{x}, \tag{1}$$

where i refers to the indices of the first and last slices, Ω_i is the 2D image domain. The function $u_i : \Omega_i \rightarrow [0, 1]$ encodes the resulting 2D segmentation for each slice, where $u_i > \frac{1}{2}$ is foreground and $u_i \leq \frac{1}{2}$ is background. The function $g_i : \Omega_i \rightarrow [0, 1]$ encodes strong image edges as small values, and the function $f_i : \Omega_i \rightarrow (-\infty, \infty)$ is defined according to the user-provided scribble annotations, where we set $f_i := -\infty$ for foreground scribbles, $f_i := \infty$ for background scribbles, and $f_i := 0$ otherwise. Minimizing (1) results in 2D segmentations that respect the user-provided scribbles annotations and align with strong image edges.

The foreground and background regions of these 2D segmentations define hard foreground and background constraints on our 3D segmentation, respectively.

Computing an Optimal Volumetric Pathway. Once we have obtained hard constraints on the first and last slices of our image stack, we generate soft constraints on all the other slices by automatically propagating the previously computed 2D segmentations through the stack. One way to accomplish this would be to advect each foreground pixel in each 2D segmentation through the image stack according to the dense optical flow between image slices. However, this approach is unreliable since small errors in the pairwise optical flow between images accumulate quickly, as noted previously by Pan *et al.* [16].

Instead, we define an optimal volumetric pathway through the image stack that connects the previously computed 2D segmentations and encloses the pixels that are most likely to belong to the neural process of interest. In this formulation, the optimal volumetric pathway is given by the interior level sets of a *cost volume* that encodes the probability of each pixel in the image stack belonging to the neural process of interest.

We define the cost of each pixel $\mathbf{p}$ in the cost volume as the length of the shortest path that connects the previously computed 2D segmentations via $\mathbf{p}$, as described in Fig. 3. We compute the length of each path through the image stack as a function of the dense optical flow between image slices as follows. For the pixels $\mathbf{p} \in \Omega_i$, $\mathbf{q} \in \Omega_{i+1}$, and the optical flow vector $v(\mathbf{p})$, we define the length from $\mathbf{p}$ to $\mathbf{q}$ as $d(\mathbf{p}, \mathbf{q}) = |\mathbf{p} \ + v(\mathbf{p}) - \mathbf{q}|$.

To compute our cost volume, we find the minimum distances from each pixel to the 2D segmentations on the first and and last image slices in two distinct

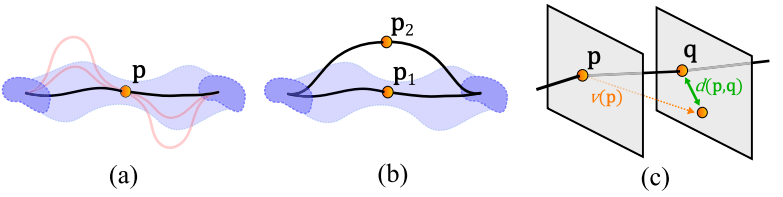


Fig. 3. Computing the cost volume. There are many possible paths through the image stack (shown in red) that connect the 2D segmentations on the first and last slices (shown in dark blue) via $\mathbf{p}$, but there is only one shortest path (shown in black); we set the cost of each pixel in the cost volume to be the length of this path (a). For example, $\mathbf{p}_2$ will be assigned a higher cost than $\mathbf{p}_1$, since the length of its shortest path is longer; this means $\mathbf{p}_2$ is less likely to belong to the neural process of interest than $\mathbf{p}_1$ (b). When computing the length of each path, we model the distance (shown in green) between pixels on adjacent slices as a function of the dense optical flow vectors (shown in orange) between the pixels; in this formulation, paths that agree strongly with the dense optical flow field have very short lengths (c).

passes, using the dynamic programming algorithm proposed by Pan et al. [16]. We set the cost of each pixel to be the sum of both distances, as proposed by Jeong *et al.* [9]. We compute dense optical flow using use the open-source implementation of Farnebäck's algorithm [5] in The OpenCV Library [3].

The interior level sets of our cost volume define soft foreground constraints on our 3D segmentation.

Computing the 3D Segmentation. Once we have obtained hard constraints on the first and last slices of the image stack, and soft constraints on all other slices, we obtain the final 3D segmentation by globally minimizing the following variational segmentation energy [17]:

$$\underset{u}{\operatorname{argmin}} \int_{\mathbf{x}\in\Omega} \left(g|\nabla^{\mathrm{xy}} u| + |\nabla^{\mathrm{z}} u| + f u\right) d\mathbf{x}, \tag{2}$$

where Ω is the 3D image domain corresponding to the entire image stack, and $\nabla^{\mathrm{xy}} u$ and $\nabla^{\mathrm{z}} u$ are the in-plane and out-of-plane gradients of u, respectively. As in (1), the function u encodes the resulting segmentation, the function g encodes strong in-plane image edges as small values, and the function f encodes constraints on the segmentation. Using the hard and soft constraints computed in the previous sections, we set $f := -\infty$ for hard foreground constraints, $f := \infty$ for hard background constraints, $f := c$ for some scalar value $c \in \mathbb{R}^-$ for soft foreground constraints, $f := c$ for some scalar value $c \in \mathbb{R}^+$ for soft background constraints, and $f := 0$ otherwise. Minimizing (2) results in a smooth 3D segmentation that respects the previously computed constraints, follows the neural process of interest, and aligns with strong in-plane image edges.

Minimizing the 2D and 3D Segmentation Energies. We compute the global minimum of (1) and (2) using the iterative parallel algorithm proposed by Unger *et al.* [17, 18]. We begin by reformulating these equations as optimization

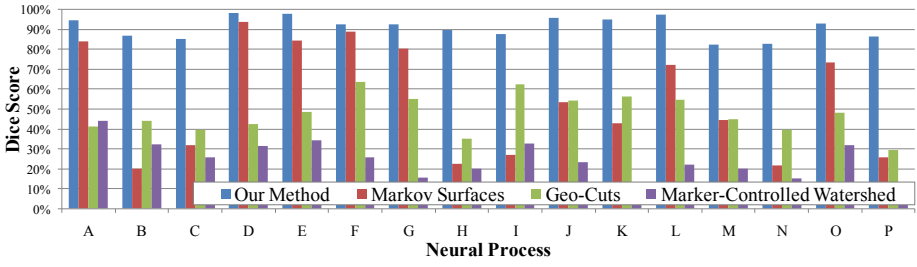


Fig. 4. Accuracy of our method, Markov Surfaces [16], Geo-Cuts [2], and Marker-Controlled Watersheds [6] while segmenting 16 neural processes in an annotated $1024 \times 1024 \times 50$ mouse hippocampus EM image stack.

problems of two variables to preserve their differentiability. In this two-variable formulation, (1) and (2) become strictly convex, so we use a projected gradient descent strategy to obtain a globally optimal solution for the segmentation variable u.

3 Results and Discussion

We evaluated our method, Markov Surfaces [16], Geo-Cuts [2], and Marker-Controlled Watersheds [6] by segmenting 16 neural processes in a $1024 \times 1024 \times 50$ mouse hippocampus image stack for which the ground truth classification of each neural process was known.

We implemented our method in CUDA and C++ on a PC with an Intel Xeon 3 GHz CPU (12GB of memory), and an NVIDIA GTX 480 graphics processor (1.5GB of memory). In all cases, computing each 2D segmentation took at most 3 seconds, computing the cost volume took at most 5 seconds, and computing the final 3D segmentation took at most 10 seconds. Total segmentation times, including all user interaction and computation time, varied between 45 and 70 seconds, with an average of 50 seconds per neural process.

Fig. 4 shows the Dice Scores[1] of all the methods used and neural processes segmented in our evaluation. On average, our method is 68% more accurate than Markov Surfaces [16], 91% more accurate than Geo-Cuts [2], and 263% more accurate than Marker-Controlled Watersheds [6]. Fig. 5 shows the segmentation results for one neural process from Fig. 4 across several 2D image slices.

As indicated in Fig. 5, Marker-Controlled Watersheds tended to under-segment neural processes, since it places segmentation boundaries at local intensity maxima. Geo-Cuts performed reasonably well on slices containing scribble annotations, but generally tended to miss the correct segmentation on other slices, since it does not propagate segmentation constraints through the image stack. Markov Surfaces performed well in regions with well-delineated boundaries, but tended to diverge heavily from image features, since it relies on Bézier interpolation

[1] $\frac{2|G \cap S|}{|G|+|S|}$ for a ground truth set of pixels G and a set of segmented pixels S.

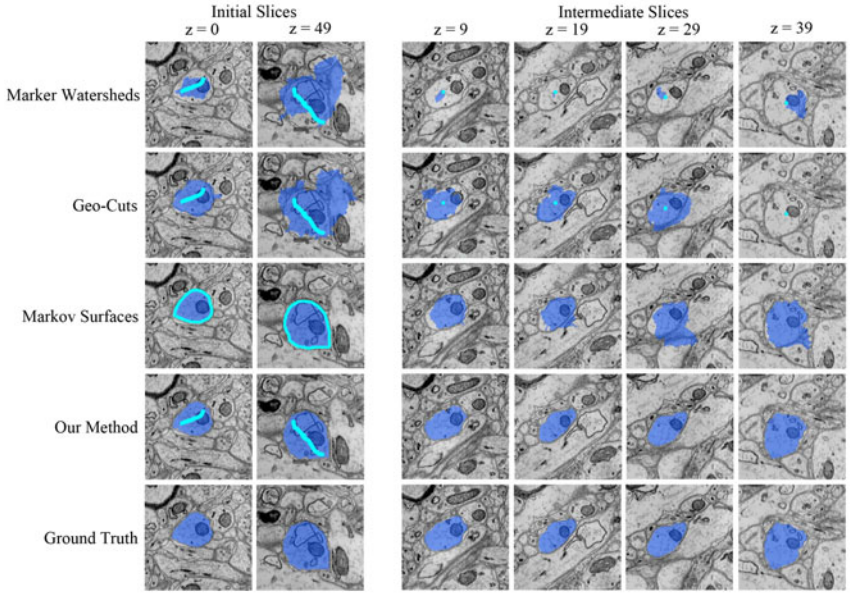


Fig. 5. Segmentation results from our method, Markov Surfaces [16], Geo-Cuts [2], and Marker-Controlled Watersheds [6] on various slices of a $1024 \times 1024 \times 50$ mouse hippocampus EM image stack. Bright blue regions indicate user-provided annotations used to initialize the algorithm, dark blue regions indicate the resulting segmentations.

to produce smooth surfaces and does not enforce 3D geometric constraints on the segmentation. Our method outperformed these methods, due to its ability to robustly propagate segmentation constraints through the image stack and optimally enforce 3D smoothness on the segmentation.

Acknowledgements. This work was partially supported by the Intel Science and Technology Center for Visual Computing, NVIDIA, the National Institute of Health under Grant No. 1P30NS062685-01 and R01 NS020364-23, the Gatsby Charitable Foundation under Grant GAT3036-Connectomic Consortium, and the National Science Foundation under Grant No. PHY-0835713. We also wish to thank Yongsheng Pan for providing us with the Markov Surfaces source code, and Manhee Lee for helping us to collect ground truth data.

References

1. Andres, B., Köthe, U., Helmstaedter, M., Denk, W., Hamprecht, F.A.: Segmentation of SBFSEM volume data of neural tissue by hierarchical classification. In: Rigoll, G. (ed.) DAGM 2008. LNCS, vol. 5096, pp. 142–152. Springer, Heidelberg (2008)
2. Boykov, Y., Funka-Lea, G.: Graph cuts and efficient N-D image segmentation. Int. J. Comp. Vis. 70(2), 109–131 (2006)

3. Bradski, G.: The OpenCV Library. Dr. Dobb's Journal of Software Tools (2000)
4. Chklovskii, D.B., Vitaladevuni, S., Scheffer, L.K.: Semi-Automated reconstruction of neural circuits using electron microscopy. Curr. Opin. Neurobiol. 20(5), 667–675 (2010)
5. Farnebäck, G.: Polynomial Expansion for Orientation and Motion Estimation. Ph.D. thesis, Linköping University, Sweden (2002)
6. Gonzalez, R.C., Woods, R.E.: Digital Image Processing, 3rd edn. Prentice-Hall, Inc., Englewood Cliffs (2006)
7. Jain, V., Bollmann, B., Richardson, M., Berger, D.R., Helmstaedter, M.N., Briggman, K.L., Denk, W., Bowden, J.B., Mendenhall, J.M., Abraham, W.C., Harris, K.M., Kasthuri, N., Hayworth, K.J., Schalek, R., Tapia, J.C., Lichtman, J.W., Seung, H.S.: Boundary learning by optimization with topological constraints. In: IEEE CVPR, pp. 2488–2495 (2010)
8. Jeong, W.K., Beyer, J., Hadwiger, M., Vazquez-Reina, A., Pfister, H., Whitaker, R.T.: Scalable and interactive segmentation and visualization of neural processes in EM datasets. IEEE Trans. Vis. Comp. Graph. 15(6), 1505–1514 (2009)
9. Jeong, W.K., Fletcher, P.T., Tao, R., Whitaker, R.T.: Interactive visualization of volumetric white matter connectivity in DT-MRI using a parallel-hardware Hamilton-Jacobi solver. IEEE Trans. Vis. Comp. Graph. 13(6), 1480–1487 (2007)
10. Jurrus, E., Whitaker, R., Jones, B.W., Marc, R., Tasdizen, T.: An optimal-path approach for neural circuit reconstruction. In: ISBI, pp. 1609–1612 (2008)
11. Kaynig, V., Fuchs, T.J., Buhmann, J.M.: Geometrical consistent 3D tracing of neuronal processes in ssTEM data. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6362, pp. 209–216. Springer, Heidelberg (2010)
12. Kaynig, V., Fuchs, T., Buhmann, J.M.: Neuron geometry extraction by perceptual grouping in ssTEM images. In: IEEE CVPR, pp. 2902–2909 (2010)
13. Lucchi, A., Smith, K., Achanta, R., Lepetit, V., Fua, P.: A fully automated approach to segmentation of irregularly shaped cellular structures in EM images. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6362, pp. 463–471. Springer, Heidelberg (2010)
14. Macke, J.H., Maack, N., Gupta, R., Denk, W., Schlkopf, B., Borst, A.: Contour-propagation algorithms for semi-automated reconstruction of neural processes. J. Neurosci. Methods 167(2), 349–357 (2008)
15. Mishchenko, Y.: Automation of 3D reconstruction of neural tissue from large volume of conventional serial section transmission electron micrographs. J. Neurosci. Methods 176(2), 276–289 (2009)
16. Pan, Y., Jeong, W.K., Whitaker, R.T.: Markov surfaces: A probabilistic framework for user-assisted three-dimensional image segmentation. In: PMMIA, pp. 57–68 (2009)
17. Unger, M., Mauthner, T., Pock, T., Bischof, H.: Tracking as segmentation of spatial-temporal volumes by anisotropic weighted TV. In: EMMCVPR, pp. 193–206 (2008)
18. Unger, M., Pock, T., Bischof, H.: Continuous globally optimal image segmentation with local constraints. In: CVWW (2008)
19. Vázquez-Reina, A., Miller, E., Pfister, H.: Multiphase geometric couplings for the segmentation of neural processes. In: IEEE CVPR, pp. 2020–2027 (2009)
20. Venkataraju, K.U., Paiva, A.R.C., Jurrus, E., Tasdizen, T.: Automatic markup of neural cell membranes using boosted decision stumps. In: ISBI, pp. 1039–1042 (2009)

An Expectation Maximization Based Method for Subcellular Particle Tracking Using Multi-angle TIRF Microscopy*

Liang Liang, Hongying Shen, Pietro De Camilli,
Derek K. Toomre, and James S. Duncan

Yale University, New Haven, CT 06520, USA
liang.liang@yale.edu

Abstract. Multi-angle total internal reflection fluorescence microscopy (MA-TIRFM) is a new generation of TIRF microscopy to study cellular processes near dorsal cell membrane in 4 dimensions (3D+t). To perform quantitative analysis using MA-TIRFM, it is necessary to track subcellular particles in these processes. In this paper, we propose a method based on a MAP framework for automatic particle tracking and apply it to track clathrin coated pits (CCPs). The expectation maximization (EM) algorithm is employed to solve the MAP problem. To provide the initial estimations for the EM algorithm, we develop a forward filter based on the most probable trajectory (MPT) filter. Multiple linear models are used to model particle dynamics. For CCP tracking, we use two linear models to describe constrained Brownian motion and fluorophore variation according to CCP properties. The tracking method is evaluated on synthetic data and results show that it has high accuracy. The result on real data confirmed by human expert cell biologists is also presented.

1 Introduction

With the assistance of live cell microscopy, especially TIRFM due to its unique illumination mechanism, biologists can investigate the underlying mechanisms of cellular processes taking place near the cell cortex, and understand some disease processes. The cell cortex is the cell's portal for intercellular communication and integrates signaling and cytoskeleton remodeling to control exo- and endocytosis. Here, we focus on clathrin mediated endocytosis (CME), which is an essential cellular process [1] that cells use for the selective internalization of surface molecules and of extracellular material. The study of CME has profound implications in neuroscience and virology. For instance, CME is the major route for synaptic vesicle recycling in neurons critical for synaptic transmission [1], and it is also one of the pathways through which viruses enter cells [2]. To study the process, the membrane associated protein complexes, namely clathrin coated pits (CCPs), are usually imaged by TIRFM. However conventional TIRFM can not provide

* This work was supported in part by the Keck Foundation and NIH grants (R37NS036251 and P30-DK45735 to P.D.C.).

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 629–636, 2011.

accurate information about z-positions and relative fluorophore amounts of individual particles. For MA-TIRFM [3], a set of images are obtained by quickly varying the incident angle, which can reveal the 3D information of the particles.

Some particle tracking methods for biological applications have been proposed in the literature [4,5,6,7]. A joint probabilistic data association (JPDA) filter based method [4] is proposed to track microtubule tips whose trajectories often cross over each other. An interacting multiple model (IMM) filter based method [5] is applied to track quantum dots with changing motion patterns. This method assumes that one motion pattern can be well described by one linear model, which is not necessarily true for our application. Recently, we proposed a 2D particle tracking method using TIRFM [6], which does better than the method in [7] for CCP tracking by incorporating information on the properties of CCPs. However, it does not consider the uncertainties from the feature detection stage, which may lead to relatively low estimation accuracy.

In this paper, we present an automatic tracking method based on MAP-Bayesian analysis to find the most probable trajectories of individual particles in 3D+t using MA-TIRFM. We adopt the basic ideas of probabilistic data association and multiple model method [8], and assume that particle dynamics at each time can be described by a set of models with a certain probability distribution. In section 2, we present the tracking framework and describe some details. In section 3, we report the evaluation results on synthetic datasets with different SNRs, and also show the result on real data.

2 Method

2.1 The Tracking Framework

Let $\mathbf{I}_t$ be the MA-TIRFM image stack acquired at time t (stack index), consisting of 2D images taken at different angles. Let $\mathbf{X}_t$ be the joint state of all particles at time t, assembled from each particle's state $X_t^{(i)}$. Let $\mathbf{D}_t$ be the observation/measurement set, and $D_t^{(j)}$ be the observation of a single particle. The goal is to find particle states that maximize the posterior probability:

$$\{\widehat{\mathbf{X}}_t\}_{t=1}^{t_{max}} = \arg \max_{\{\mathbf{X}_t\}_{t=1}^{t_{max}}} \log p\left(\{\mathbf{X}_t\}_{t=1}^{t_{max}} \left| \{\mathbf{I}_t\}_{t=1}^{t_{max}}\right.\right) \tag{1}$$

Here, $\{\mathbf{X}_t\}_{t=1}^{t_{max}} = \{\mathbf{X}_1, ..., \mathbf{X}_{t_{max}}\}$. Since it is difficult to find the optimal solution directly, a practical way is to run detection and tracking iteratively:

$$\text{Detection}: \{\widehat{\mathbf{D}}_\tau\}_{\tau=1}^{t} = \arg \max_{\{\mathbf{D}_\tau\}_{\tau=1}^{t}} \log p\left(\{\mathbf{D}_\tau\}_{\tau=1}^{t} \left| \{\widehat{\mathbf{X}}_\tau, \mathbf{I}_\tau\}_{\tau=1}^{t}\right.\right) \tag{2}$$

$$\text{Tracking}: \{\widehat{\mathbf{X}}_\tau\}_{\tau=1}^{t} = \arg \max_{\{\mathbf{X}_\tau\}_{\tau=1}^{t}} \log p\left(\{\mathbf{X}_\tau\}_{\tau=1}^{t} \left| \{\widehat{\mathbf{D}}_\tau\}_{\tau=1}^{t}\right.\right) \tag{3}$$

Detection is to find the most likely features of the particles given the image data and previously estimated states (if available). We have developed an algorithm

for 3D particle feature estimation (3D positions and relative amount of fluorophores) using MA-TIRFM similar to the method in [3]. Tracking is to find the most probable states/trajectories of the particles given the observation data.

To solve the MAP problem (Eq.(3)), we develop an expectation maximization (EM) based algorithm. We also develop a forward filter based on the MPT filter [8,9] and the idea of probabilistic data association [4] to provide initial estimations for the EM algorithm. The models of particle dynamics are described in section 2.2. The EM algorithm and the forward filter are described in section 2.3 and 2.4 respectively. The tracking flow is presented in section 2.5.

2.2 The Particle Dynamics Models

We assume particle dynamics/modes can be modeled using linear state space models [8] with a certain probability distribution at each time. $\pi(k,n)$ is the prior transition probability for each particle to switch from mode (model) n to k when the current mode (model) is n, and $\sum_k \pi(k,n) = 1$. Each model k has a posterior probability $\phi_t^{(i,k)}$ for each particle i, and $\sum_k \phi_t^{(i,k)} = 1$. Each model is given as

$$X_t^{(i)} = F_t^{(k)} X_{t-1}^{(i)} + U_t^{(i,k)} + W_t^{(k)} \tag{4}$$

$$D_t^{(j)} = H X_t^{(i)} + V_t^{(j)} \tag{5}$$

Here, $F_t^{(k)}$ is the state transition matrix. $U_t^{(i,k)}$ is the external input that we use to impose constraints. $W_t^{(k)}$ is the process noise with covariance matrix $Q_t^{(k)}$ learned from training data. H is a constant observation matrix. $V_t^{(j)}$ is the observation noise with covariance matrix $R_t^{(j)}$ provided by the detection module. Each of these noise sources is assumed to be Gaussian and independent.

We define the state of each CCP i at time t as $X_t^{(i)} = [x_t, y_t, z_t, c_t, \dot{c}_t]'$. $[x_t, y_t, z_t]$ is the position of the particle. c_t is the relative amount of fluorophores in the particle, and $\dot{c}_t$ is the rate of change of fluorophores over time. We propose to use two linear state space models. For particle motion, the first model describes it as free Brownian motion because the motion is indeed random, and the second model describes it as confined motion because each CCP is linked to the plasma membrane through its neck [1,6]. For fluorophore variation in each CCP, both models describe it as a linear process. The parameters are given by

$$F_t^{(1)} = \begin{bmatrix} 1&0&0&0&0 \\ 0&1&0&0&0 \\ 0&0&1&0&0 \\ 0&0&0&1&1 \\ 0&0&0&0&1 \end{bmatrix}, \; F_t^{(2)} = \begin{bmatrix} 0&0&0&0&0 \\ 0&0&0&0&0 \\ 0&0&0&0&0 \\ 0&0&0&1&1 \\ 0&0&0&0&1 \end{bmatrix}, \; U_t^{(i,2)} = \begin{bmatrix} \bar{x}_{t-1} \\ \bar{y}_{t-1} \\ \bar{z}_{t-1} \\ 0 \\ 0 \end{bmatrix}, \; H = \begin{bmatrix} 1&0&0&0&0 \\ 0&1&0&0&0 \\ 0&0&1&0&0 \\ 0&0&0&1&0 \end{bmatrix}$$

$U_t^{(i,1)}$ is a zero vector for the first model, and $U_t^{(i,2)}$ is the expected position of the particle determined by its neck joint. $[\bar{x}_t, \bar{y}_t, \bar{z}_t]$ is the time-average position, i.e., $\bar{x}_t = \frac{1}{t-t_0+1} \sum_{\tau=t_0}^{t} x_\tau$, and t_0 is the starting-time of the trajectory. Since each model reveals just half truth, we set the prior transition probabilities to be equal. The initial model probabilities are given as $\phi_1^{(i,1)} = 1$ and $\phi_1^{(i,2)} = 0$.

2.3 The EM Algorithm

In the literature [8], some EM-based algorithms have already been proposed for general multiple target tracking, in which either the model probabilities or the data association probabilities are hidden variables. Here, we treat all of them as hidden variables, and combine deterministic annealing with the EM algorithm, and make some simplifications.

The lower bound $\mathcal{B}$ of the objective function (Eq.(3)) is constructed as:

$$\mathcal{B}\left(\{\mathbf{X}_\tau\}_{\tau=1}^t;\{\widetilde{\mathbf{X}}_\tau\}_{\tau=1}^t\right)=\mathrm{E}\left[\log p\left(\{\mathbf{D}_\tau,\mathbf{X}_\tau,\mathbf{A}_\tau,\mathbf{\Phi}_\tau\}_{\tau=1}^t\right)\middle|\{\mathbf{D}_\tau,\widetilde{\mathbf{X}}_\tau\}_{\tau=1}^t\right] +\mathcal{H}\left(p\left(\{\mathbf{A}_\tau,\mathbf{\Phi}_\tau\}_{\tau=1}^t\middle|\{\mathbf{D}_\tau,\widetilde{\mathbf{X}}_\tau\}_{\tau=1}^t\right)\right) \tag{6}$$

Here, $\{\widetilde{\mathbf{X}}_\tau\}_{\tau=1}^t$ is the estimated states from the previous iteration.$\mathcal{H}(p)$ is the entropy of the probability distribution p. $\mathbf{A}_\tau$ is the binary matrix of data association. $\mathbf{A}_\tau(i,j)=1$ is the event that the observation $D_\tau^{(j)}$ is from the particle i at time τ, and $\mathbf{A}_\tau(i,0)=1$ is the event that the particle i disappears, and $\mathbf{A}_\tau(0,j)=1$ is the event that a new particle appears with observation $D_\tau^{(j)}$. Let $a_\tau^{(i,j)}$ be the data association probability, $a_\tau^{(i,j)}=p\left(\mathbf{A}_\tau(i,j)=1\middle|\{\mathbf{D}_\tau,\widetilde{\mathbf{X}}_\tau\}_{\tau=1}^t\right)$ with $\sum_{i>0}a_\tau^{(i,j)}=1$ and $\sum_{j>0}a_\tau^{(i,j)}=1$ assuming one-to-one correspondences. $\mathbf{\Phi}_\tau$ is the binary matrix of model assignment. $\mathbf{\Phi}_\tau(i,k)=1$ means the k^{th} model matches the dynamics of the particle i at time τ. The model probability $\phi_\tau^{(i,k)}$ is defined as $\phi_\tau^{(i,k)}=p\left(\mathbf{\Phi}_\tau(i,k)=1\middle|\{\mathbf{D}_\tau,\widetilde{\mathbf{X}}_\tau\}_{\tau=1}^t\right)$ with $\sum_k\phi_\tau^{(i,k)}=1$.

By using Eq.(4)-(6) and assuming that $\mathbf{A}_\tau$ and $\mathbf{\Phi}_\tau$ are statistically independent given the states and observation data, and ignoring constant terms, and combining deterministic annealing method, we obtain the expression of $\mathcal{B}$:

$$\begin{aligned}\mathcal{B}=&-\frac{1}{2}\sum_{\tau=1}^t\sum_{i,j,k}a_\tau^{(i,j)}\phi_\tau^{(i,k)}\left(\left\|X_\tau^{(i)}-F_\tau^{(k)}X_{\tau-1}^{(i)}-U_\tau^{(i,k)}\right\|_{Q_\tau^{(k)}}^2+\log(|Q_\tau^{(k)}|)\right.\\&\left.+\left\|D_\tau^{(j)}-HX_\tau^{(i)}\right\|_{R_\tau^{(j)}}^2+\log(|R_\tau^{(j)}|)\right)-T\sum_{\tau=1}^t\sum_{i,j}a_\tau^{(i,j)}\log\left(a_\tau^{(i,j)}\right)\\&-\sum_{\tau=1}^t\sum_{i,k}\phi_\tau^{(i,k)}\log\left(\phi_\tau^{(i,k)}\right)+\sum_{\tau=1}^t\sum_{i,k}\phi_\tau^{(i,k)}\sum_n\phi_{\tau-1}^{(i,n)}\log\pi(k,n)\end{aligned} \tag{7}$$

Here, the vector norm is defined as $\|V\|_L^2=V'L^{-1}V$. T is the annealing temperature. $D_\tau^{(0)}$ is virtual observation specified for each particle similar to those in [6], which is commonly used to handle particle disappearing.

The exact formula for $a_\tau^{(i,j)}$ is very complicated for implementation, and therefore we make some approximations. By taking the derivative of $\mathcal{B}$ with respect to $a_\tau^{(i,j)}$ at $\{\mathbf{X}_\tau\}_{\tau=1}^t=\{\widetilde{\mathbf{X}}_\tau\}_{\tau=1}^t$, setting the result to zero, and making some modifications for robustness purpose, we obtain:

$$a_\tau^{(i,j)} \propto \exp\left[\frac{-1}{T}\sum_k \phi_\tau^{(i,k)}\left(\left\|H\left(\widetilde{X}_\tau^{(i)} - \widetilde{X}_{\tau|\tau-1}^{(i,k)}\right)\right\|^2_{\widetilde{Q}_\tau^{(k)}} + \left\|D_\tau^{(j)} - H\widetilde{X}_\tau^{(i)}\right\|^2_{R_\tau^{(j)}}\right)\right] \quad (8)$$

Here, $\widetilde{X}_{\tau|\tau-1}^{(i,k)} = F_\tau^{(k)}\widetilde{X}_{\tau-1}^{(i)} + U_\tau^{(i,k)}$, and $\widetilde{Q}_\tau^{(k)} = HQ_\tau^{(k)}H'$. The softassign algorithm [10] is applied to ensure that $\sum_{i>0} a_\tau^{(i,j)} = 1$ and $\sum_{j>0} a_\tau^{(i,j)} = 1$.

Similarly, we can obtain the update formula for $\phi_\tau^{(i,k)}$:

$$\phi_\tau^{(i,k)} \propto \exp\left[-\frac{1}{2}\left\|H\left(\widetilde{X}_\tau^{(i)} - \widetilde{X}_{\tau|\tau-1}^{(i,k)}\right)\right\|^2_{\widetilde{Q}_\tau^{(k)}} + \sum_n \phi_{\tau-1}^{(i,n)}\log\pi(k,n) + \sum_n \phi_{\tau+1}^{(i,n)}\log\pi(n,k)\right] \quad (9)$$

Therefore, in the E-step, the probabilities of models and data associations are calculated using Eq.(8)-(9). In the M-step, the particle states that maximize $\mathcal{B}$ are estimated, which is achieved by taking the derivative of $\mathcal{B}$ with respect to the states of each particle, setting the result to zero, and solving a linear system of equations. In this step, the terms containing virtual observations are ignored. After each EM iteration, the annealing temperature is reduced. Eventually each $a_\tau^{(i,j)}$ converges to 0 or 1, and one particle state has one observation.

2.4 The Forward Filter

We develop a forward filter based on the most probable trajectory (MPT) filter [8,9] that was proposed for tracking a single target with multiple models. The forward filter handles the data association problem in a manner consistent with the EM algorithm, and provides state estimation and prediction given the observation data. Following the similar derivation process and notations in [9], we obtain the prediction and update formulas for each particle's state:

$$X_{t+1|t}^{(i)} = K_{t+1}^{-1}\left[\overline{QU_t} + \overline{QF_t}S_t\left(-\overline{F'QU_t} + \overline{H'RD_t} + K_tX_{t|t-1}^{(i)}\right)\right] \quad (10)$$

$$X_{t|t}^{(i)} = \left(\overline{F_t}\right)^{-1}\left(X_{t+1|t}^{(i)} - \overline{U_t}\right) \quad (11)$$

$$S_t = \left(\overline{F'QF_t} + \overline{H'RH_t} + K_t\right)^{-1} \quad (12)$$

$$K_{t+1} = \overline{Q_t} - \overline{QF_t}S_t\overline{F'Q_t} \quad (13)$$

Here, we define $\overline{G_t} = \sum_k \phi_t^{(i,k)}G_t^{(i,k)}$ and $\overline{GQP_t} = \sum_k \phi_t^{(i,k)}G_t^{(i,k)}(Q_t^{(k)})^{-1}P_t^{(i,k)}$ for any matrix G and P, e.g., $\overline{F'QU_t} = \sum_k \phi_t^{(i,k)}(F_t^{(k)})'(Q_t^{(k)})^{-1}U_t^{(i,k)}$, and $\overline{H'RH_t} = \sum_{j>0} a_t^{(i,j)}H'(R_t^{(j)})^{-1}H$, and $\overline{H'RD_t} = \sum_{j>0} a_t^{(i,j)}H'(R_t^{(j)})^{-1}D_t^{(j)}$.

We define the data association probabilities and model probabilities as:

$$a_t^{(i,j)} \propto \exp\left[-\sum_k \phi_{t|t-1}^{(i,k)} \left\| D_t^{(j)} - HX_{t|t-1}^{(i)} \right\|^2_{\widetilde{Q}_t^{(k)}}\right] \tag{14}$$

$$\phi_t^{(i,k)} \propto \phi_{t|t-1}^{(i,k)} \exp\left[-\frac{1}{2}\sum_j a_t^{(i,j)} \left\| D_t^{(j)} - H\left(F_t^{(k)} X_{t-1|t-1}^{(i)} + U_t^{(i,k)}\right)\right\|^2_{\widetilde{Q}_t^{(k)}}\right] \tag{15}$$

$$\phi_{t|t-1}^{(i,k)} = \sum_n \pi(k,n)\, \phi_{t-1}^{(i,n)} \tag{16}$$

Here, $\sum_k \phi_t^{(i,k)} = 1$, $\sum_{i>0} a_t^{(i,j)} = 1$ and $\sum_{j>0} a_t^{(i,j)} = 1$.

2.5 The Tracking Flow

After the detection module runs on the image stacks, the tracking module starts with the forward filter to obtain state estimations in the time window $[t, t+w]$. Then, the EM algorithm runs in the time window to update the relevant estimations. Given the updated states as prior information, in the time window the detection module runs again to refine the detection results and the EM algorithm runs again to refine state estimations. Then, the forward filter is reinitialized at time $t+w$, and the window slides, and the process is repeated. During the tracking process, a track i is terminated when $a_t^{(i,0)}$ is greater than a threshold (0.9 in experiments), and new tracks are initialized if the number of observations is greater than the number of existing particles.

3 Experimental Results

3.1 Evaluation on Synthetic Data

The proposed method is evaluated on synthetic MA-TIRFM image datasets. Each dataset contains 200 image stacks, and each image stack consists of four images acquired at penetration depth 100nm, 200nm, 300nm, and 400nm. Gaussian distributions are fitted to the histogram of particle displacements and to the histogram of the deviations from the mean positions of individual particles, and therefore trajectories can be created by sampling from the distributions. Three major types of noises are considered in the simulation process, the thermal noise of laser, the Poisson (shot) noise of input photon, and the excess noise generated in the EMCCD. The SNR of each dataset can be tuned by varying the ratio of the average fluorophore amount of individual CCPs and the background fluorophore concentration. We also test an alternative method which is essentially the same as the proposed method except that it only uses the first model described in section 2.2.

The results are shown in Fig. 1(a)-(h). Tracking accuracies (defined in [6]) under different SNRs are shown in Fig.1(a). The proposed method is consistently better. The alternative method produces some long but wrong trajectories

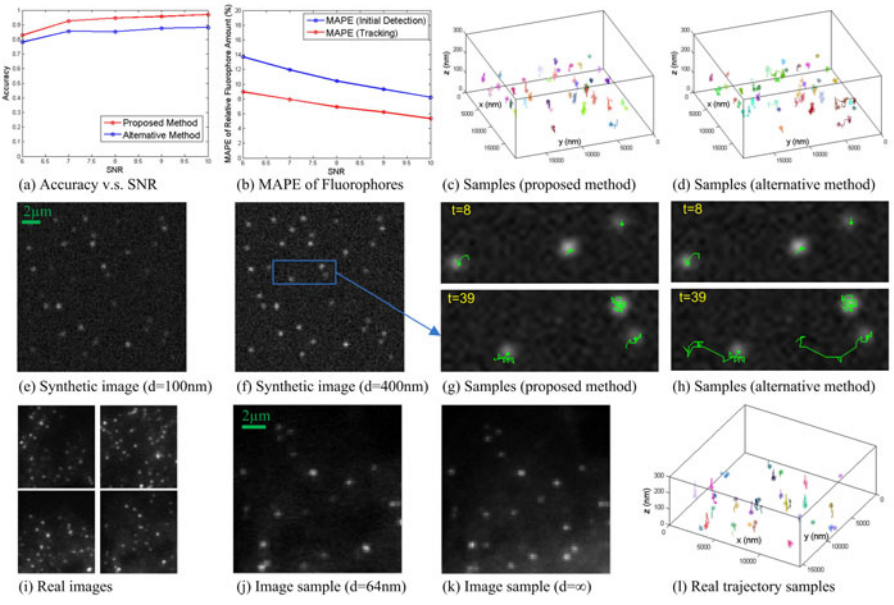


Fig. 1. (a) Accuracy-SNR curves. (b) MAPE of the estimated relative amount of fluorophores. (c)&(d) Trajectory samples in 3D. (e)&(f) Synthetic images at $d = 100nm$ and $d = 400nm$. (g)&(h) Trajectory samples in 2D. (i) Image samples of real datasets. (j)&(k) Real images acquired at $d = 64nm$ and $d = \infty$. (l) Real trajectory samples obtained by using the proposed method.

(Fig.1(h)), which happens most likely when new CCPs appear near the positions where matured CCPs disappear. The result is consistent with what is reported in [6]. In Fig.1(b), we show the mean absolute percentage error (MAPE) of the estimated relative amount of fluorophores. The MAPE calculated after tracking (using the proposed method) is smaller than the MAPE calculated after initial detection. Since the uncertainties from the detection stage are considered in the tracking framework, the method can improve feature estimation accuracy.

3.2 Evaluation on Real Data

Mouse fibroblast cells were electroporated with clathrin light chain GFP construct using the Amaxa Nucleofector method, and were plated at subconfluent densities into 35mm glass bottom dishes, and allowed to grow for 12 to 48 hours. Images were acquired by Nikon Ti-E Eclipse inverted microscope with Apo TIRF 100× oil objective lens (NA 1.49) and Andor iXon 897 EMCCD camera. For the initial evaluation, two angles were used. The first one corresponds to penetration depth $d = 64nm$, to observe the particles near the glass slide. The second one is greater than the critical angle, to observe particles in the deeper range. Although total internal reflection will not occur with the second angle, the axial profile of the light field can still be modeled as an exponential function $f(z) = I_0 e^{-z/d} = I_0$

with $d = \infty$. Mathematically, it does not make any difference to the method. Regions of interest are selected from image datasets, and the tracking results are visually inspected and confirmed by human expert cell biologists. Samples of image regions and trajectories are shown in Fig. 1 (i)-(l).

4 Conclusion

We have proposed an automatic particle tracking method based on a MAP framework and applied it to track clathrin coated pits (CCPs) in 3D+t using MA-TIRFM. We use a set of linear models to model particle dynamics, and develop the EM-based algorithm and the forward filter to solve the MAP problem. We evaluate the method using simulation with different SNRs, and also test it on real data. The results show the method is promising for the analysis of CCPs. In the near future, we'll optimize MA-TIRFM imaging parameters and obtain images with four penetration depths and two fluorescent channels.

References

1. Slepnev, V.I., De Camilli, P.: Accessory factors in clathrin-dependent synaptic vesicle endocytosis. Nature Reviews Neuroscience 1, 161–172 (2000)
2. Brandenburg, B., Zhuang, X.: Virus trafficking - learning from single-virus tracking. Nature Reviews Microbiology 5, 197–208 (2007)
3. Yang, Q., Karpikov, A., Toomre, D., Duncan, J.S.: 3D reconstruction of microtubules from multi-angle total internal reflection fluorescence microscopy using Bayesian framework. IEEE Trans. on Image Processing (2011) (in press)
4. Smal, I., Niessen, W., Meijering, E.: A new detection scheme for multiple object tracking in fluorescence microscopy by joint probabilistic data association filtering. In: IEEE Int. Symposium on Biomedical Imaging: From Nano to Macro, pp. 264–267 (2008)
5. Genovesio, A., Liedl, T., Emiliani, V., Parak, W.J., Coppey-Moisan, M., Olivo-Marin, J.-C.: Multiple particle tracking in 3-D+t microscopy: method and application to the tracking of endocytosed quantum dots. IEEE Trans. on Image Processing 15(5), 1062–1070 (2006)
6. Liang, L., Shen, H., De Camilli, P., Duncan, J.S.: Tracking clathrin coated pits with a multiple hypothesis based method. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6362, pp. 315–322. Springer, Heidelberg (2010)
7. Jaqaman, K., Loerke, D., Mettlen, M., Kuwata, H., Grinstein, S., Schmid, S.L.L., Danuser, G.: Robust single-particle tracking in live-cell time-lapse sequences. Nature methods 5, 695–702 (2008)
8. Li, X.R., Jilkov, V.P.: Survey of maneuvering target tracking. Part I. Dynamic models. IEEE Trans. on Aerospace and Electronic Systems 39(4), 1333–1364 (2003)
9. Zhang, Q.: Optimal filtering of discrete-time hybrid systems. Journal of Optimization Theory and Applications 100(1), 123–144 (1999)
10. Rangarajan, A., Chui, H., Bookstein, F.L.: The softassign procrustes matching algorithm. In: Duncan, J.S., Gindi, G. (eds.) IPMI 1997. LNCS, vol. 1230, pp. 29–42. Springer, Heidelberg (1997)

Discriminative Segmentation of Microscopic Cellular Images

Li Cheng[1], Ning Ye[1], Weimiao Yu[2], and Andre Cheah[3]

[1] BioInformatics Institute, A*STAR, Singapore
[2] Institute of Molecular and Cell Biology, A*STAR, Singapore
[3] NUHS, Singapore

Abstract. Microscopic cellular images segmentation has become an important routine procedure in modern biological research, due to the rapid advancement of fluorescence probes and robotic microscopes in recent years. In this paper we advocate a discriminative learning approach for cellular image segmentation. In particular, three new features are proposed to capture the appearance, shape and context information, respectively. Experiments are conducted on three different cellular image datasets. Despite the significant disparity among these datasets, the proposed approach is demonstrated to perform reasonably well. As expected, for a particular dataset, some features turn out to be more suitable than others. Interestingly, we observe that a further gain can often be obtained on top of using the "good" features, by also retaining those features that perform poorly. This might be due to the complementary nature of these features, as well as the capacity of our approach to better integrate and exploit different sources of information.

1 Introduction

Cellular images segmentation is an indispensable step for modern biological research, and this is greatly facilitated by the recent development of fluorescence dyes and robotic microscopes. A number of unsupervised segmentation methods [1], such as thresholding, region-growing, watershed, level-set, and edge-based methods, have been developed to address this problem in scenarios where the foreground objects and the background regions have distinct color or textural properties. Many of these methods are dedicated to specific problems where domain knowledge is heavily exploited by tuning algorithmic parameters manually. This case-by-case approach could be very tedious. On the other hand, there are many images (such as in Figure 1) that turns to be rather difficult for unsupervised approaches [1], partly due to the variations of specimen types, staining techniques, and imaging hardware. This leads to a recent development in learning-based algorithms, including support vector machines (SVMs) [10], as well as conditional random fields and variants [4]. While they demonstrate that reasonable segmentation results can be produced for some difficult cases, these methods are often still dedicated to certain type of microscopic images for specific problems.

In this paper we propose a flexible learning framework for cellular image segmentation, and we intend to show that it is possible to develop a more generic segmentation framework that works effectively over a broader spectrum of microscopic images such as those displayed in Figure 1. This is achieved by carefully integrating information

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 637–644, 2011.

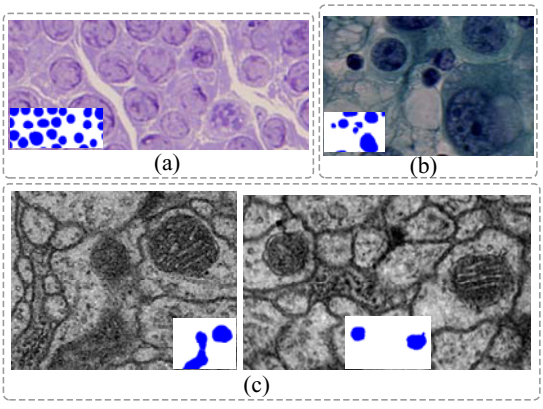


Fig. 1. Illustration of some difficult examples encountered in microscopic image segmentation. These are crop-out examples from three cellular image datasets: (a) hand, (b) serous, (c) ssTEM. Ground truth is shown in the inlet for each of the examples.

from both local and global aspects, as illustrated in Figure 2. In particular, three novel features are proposed: (1) An appearance feature that integrates both color and texture information; (2) A spoke feature that effectively encodes the shape of cellular foreground objects; (3) Meanwhile the detection score is also used to exploit the strength of object detection developed over the years in computer vision. Besides, a superpixel-based coding scheme is devised to incorporate higher-order scene context.

2 Our Approach

The flow chart of our approach is depicted in Figure 2. An image pixel is characterized by a set of features describing various local aspects in its neighborhood, such as shape, appearance, and context information. These pixel-based features are further pooled to form one vector for a superpixel or oversegment [7]. Finally, a global discriminative classifier is utilized to incorporate these superpixel-based shape, appearance, and context features to produce a segmentation prediction for the input image.

Appearance feature: Unary & **Binary Extensions of Color BoW model.** For a pixel in color images, its RGB and YUV color values are combined to form a 6D vector. For grayscale images, 1D intensity feature is used directly. As illustrated in the middle panel of Figure 3, a visual Bag-of-Words (BoW) model with K codewords is built, and these color vectors are thus mapped to the quantized space spanned by the codewords [11]. A novel appearance feature is proposed here by integrating BoW model with local neighboring information by means of unary/binary extensions: A unary extension partitions the local neighbors into disjoint parts. One scheme is to partition into concentric layers, as displayed in Figure 3(a). By pooling the codeword assignments of these features and normalizing to sum to 1, one partition is characterized by a histogram of length K codewords. A length $K \times S$ vector is thus produced by concatenating over S partitions ($S = 3$ in Figure 3(a)). Note that other partition schemes might also be possible.

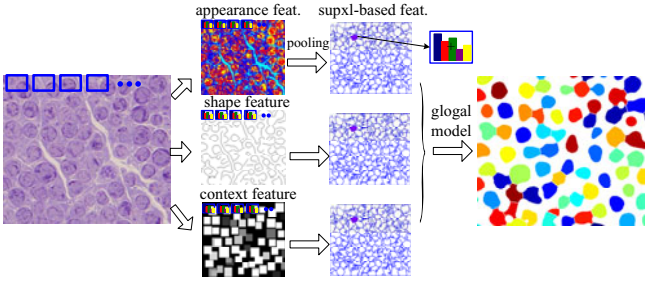


Fig. 2. An illustrative flowchart of the proposed approach. From a pixel of an input image, a set of features is extracted to capture various aspects in its neighborhood, including appearance, shape, and context information. A novel superpixel (over-segmentation) feature is devised to provide a more compact signature and to incorporate higher-order scene context. These superpixel-based features are fed to a global discriminative learning model to deliver a segmentation.

Meanwhile, a binary extension considers pairs of neighboring pixels, and similar to the concept of co-occurrence matrix, accumulates the counts into a 3D array indexed by (codewords, codewords, distance). Naively this leads to a vector of length $K \times K \times S'$, by accumulating the quantized distance of every feature pair with S' possible outcomes. Here we adopt hamming distance. In practice it is further summarized into a more compact vector representation: For a particular quantized distance, (a) a K-dim vector is extracted from the diagonal elements, and (b) a K-dim vector is obtained by summing over all the off-diagonal elements row-wise. For both cases the output vectors are normalized to sum to 1. As each case ends up giving a $K \times S'$ vector, a concatenation of both finally leads to a vector representation of length $2K \times S'$. Our final appearance feature is thus produced by concatenating both unary and binary extensions. In this paper, we fix $K = 100$, $S = 3$, and $S' = 3$.

Shape feature: Multi-scale Spoke Feature. as illustrated in Figure 4(a), for any location in an image, its spokes are equally sampled in angular space and each reach out until an edge is met. Determined by its local convexity (i.e. the orientation of its signed curvature), the spoke will contribute to one of the three bins: +, 0, and -, that encode the local shape as being convex, undecided, or concave, respectively. Therefore, the spoke feature essentially encodes the local shape information from the direct object boundaries surrounding this location, while being invariant to rigid transformation. As cellular objects often possess convex shapes, this feature ideally provide sufficient discrimination power to differentiate a location inside a cellular object from being outside. On the other hand, a single edge map usually does not faithfully retain object boundaries of the image. To address this issue, we use Canny edge detector together with its Gaussian smooth kernels at multiple scales. This gives rise to the multi-scale feature in Figure 4(c). In practice, for a pixel, 9 spokes are used and the number of scales is set to 5. The elements in the histogram vector of each scale are also normalized to sum to 1.

Context feature: Detection Score BoW model. Object detection is usually regarded as a separate problem from image segmentation, and thus dealt with by substantially

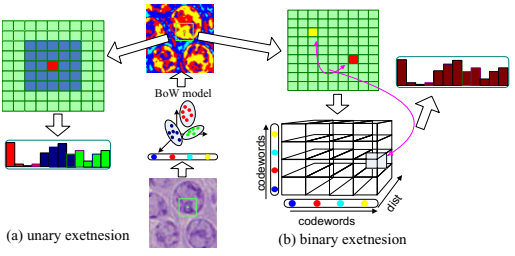


Fig. 3. Unary and binary extensions. (a) A unary feature extension partitions the window into concentric layers. By pooling the pixels' codeword assignments in the BoW model (length K) within each partition, a $K \times S$ length vector is produced as a concatenation of S histograms. Each histogram of length K comes from one partition. (b) A binary feature extension. Each pair of pixels in the window is used to accumulate the counts in a 3-dimensional array. Naively this leads to a vector of length $K \times K \times S'$, for a quantized distance of S' possible outcomes. In practice it is further summarized into a more compact vector of length $2K \times S'$. Then unary and binary extensions are concatenated together to form an appearance feature.

different techniques. Nevertheless, detection outputs possesses important information about the locations and sizes of the foreground objects that can be utilized to help segmentation. In addition, as generated through top-down schemes, the detection scores carries context information over to pixel level. Here we adopt a dedicated mixture model-based object detector [6]. The bounding box detections are overlayed onto a two dimensional space with each assigning its detection score. This is treated as a separate channel of the input image. Then the context feature is produced through a BoW model, similar to that of the color BoW model as previously shown in the middle panel of Figure 3. Here the number of bins in the BoW model is set to 4.

Superpixel-based Feature. An image is usually represented as a two-dimensional lattice where each node corresponds to a pixel. However a pixel by itself contains limited information. Alternatively, an image can be expressed as a general planar graph, and each node is now a superpixel or oversegmentation [7]containing a set of nearby pixels, usually obtained using an unsupervised segmentation. As depicted in Figure 2, for pixels within a superpixel, their features are pooled to form a higher-order feature vector that describes the entire superpixel. Similarly, the output vectors are each normalized to sum to 1. While providing a more compact feature representation, this superpixel-based feature is also able to capture higher-order scene context.

Global Model and Postprocessing. The discriminative learning method we have utilized is a Structured Support Vector Machine (SSVM) [12], where the optimal assignment problem is solved by the graph-cuts algorithm [2] that incorporates both node and edge energies to ensure local and global compatibilities. The proposed features are concatenated as a long feature vector used for the node energy; While our edge energy adopts a simple Ising model [2]. Our postprocessing step utilizes distance transform and generalized Voronoi diagram [13], to remove tiny segments and separate those lightly touched objects.

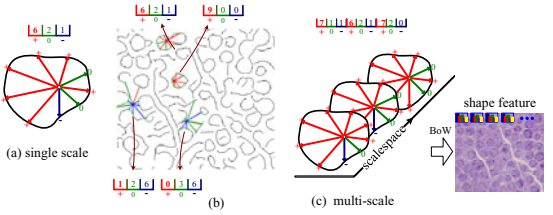


Fig. 4. (a) An illustration of the proposed *spoke feature*, (b) its usages in an edge map of the image in Figure 2 on four image locations, and (c) a *multi-scale spoke feature*. Spokes are equally sampled in angular space, and are further mapped into a histogram vector of three bins marked by +, 0, and -, which denote locally convex, undecided, and concave, respectively. To alleviate the issues introduced by edge detection, edge maps are extracted at multiple scales, and the associated histograms are concatenated to form a multi-scale spoke feature. A common procedure in edge (e.g. Canny) detectors is to convolve raw image with Gaussian kernel of certain width, which is regarded as selecting a scale-space [9]. Multi-scale here refers to applying kernel of multiple widths, leading to multiple edge maps. For a fixed image location, a multi-scale spoke feature is obtained by concatenating the spoke feature vectors obtained over scales.

3 Experiments

Image Datasets. Three image datasets are used during the experiments, where for each dataset, half of the images are used for training and the rest images are retained for testing purpose. The *hand dataset* contains images of hand nerve endings harvested from fresh frozen adult human cadavers. They are preserved in gluteradehyde and refridgerated, then embedded in liquid wax to form a block to facilitate microtome sectioning of the specimen perpendicular to the longitudinal axis of the axons. They are then stained with methylene blue and photographed using a light microscope to facilitate the process of histomorphometry. They are further partitioned into 24 smaller images of similar sizes, with ground-truth provided for nerve endings. To study the drosophila first instar larva ventral nerve cord (VNC), the *ssTEM dataset* is generated, which has 30 images[3][1] taken from a serial section Transmission Electron Microscopy (ssTEM), with image resolution 4x4 nm/pixel. Ground-truth annotations are provided for mitochondria. Finally, the *serous dataset* [8] contains 10 microscopic images[2] from serous cytology. Ground-truth annotations are provided for cell nuclei.

Performance Evaluation. Performance of a cellular segmentation method is often quantitatively assessed by two types of metrics: those of pixel-based and those of object-based. We follow the PASCAL VOC evaluation criteria [5]: For pixel-level, we directly adopt the criteria of its image segmentation task [5]. The metric for object-level evaluation is an adaptation of the criteria used in its object detection task [5]. For *Pixel-based*

[1] The dataset is downloaded from
`http://www.ini.uzh.ch/~acardona/data/tifs.tar.bz2`
[2] The dataset is downloaded from
`http://users.info.unicaen.fr/~lezoray/databases/SerousDatabase.zip`

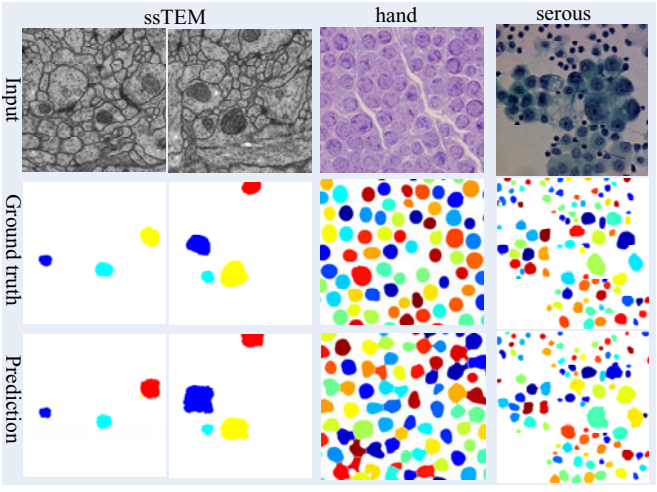


Fig. 5. Sample experimental results on the three datasets

Evaluation , a common accuracy measures the average percentage of pixels being correctly classified for both foreground and background classes. This metric however can be misleading when class distribution is unbalanced, e.g. when the dataset contains fewer foreground object pixels and a larger percentage of background pixels (the ssTEM dataset). To rectify this issue, the PASCAL image segmentation task advocates to compute the accuracy as (Eq.(4) of [5]) $\frac{\mathrm{TP}}{\mathrm{TP+FN+FP}}$[3]. For *object-level Evaluation* , object-based image segmentation can be considered as a special object detection task, where in addition to location and scale, it also demands the detailed shape of a foreground object. Following the object detection task of PASCAL [5], we use an intersection/union ratio to determine a correct object-level match: Given a pair of objects consisting of a prediction area O_{p} and a ground-truth O_{gt}, there exists a match if the overlap ratio, $\frac{\mathrm{area}(O_{\mathrm{p}} \cap O_{\mathrm{gt}})}{\mathrm{area}(O_{\mathrm{p}} \cup O_{\mathrm{gt}})}$ exceeds a threshold t, where $\cap$ denotes the intersection, and $\cup$ the union. t is set to 0.5 as in [5]. Similarly we define object-level accuracy as $\frac{\mathrm{TP}}{\mathrm{TP+FN+FP}}$.

Experiments. Throughout the experiments, the unsupervised method of [7] is used to partition an image into superpixels, and the C value of the linear SSVM [12] is fixed to 100.

As expected, a multi-scale shape feature leads to improved performance, when comparing to its single-scale counterpart. This is demonstrated in hand dataset, where the pixel- (and object-) level accuracy is around 68% vs. 51% (and 76% vs. 73%), when a multi-scale shape feature is compared to a single-scale one. We also observe that *no single* feature excels in all datasets. For example, the appearance feature dominates the performance for the serous dataset, while it works less well in the hand dataset, and leads to the worst results for the ssTEM dataset. Meanwhile, it is mostly preferable to

[3] TP, TN, FP, and FN refer to True Positive, True Negative, False Positive and False Negative, respectively.

Table 1. Comparisons of pixel- & object- level accuracies. Here 'three features' refers to the full version of our proposed method; 'appearance', 'shape', and 'context' are variants where only one type of feature is used; Meanwhile, 'dedicated unsupervised seg.' refers to the segmentation method in [13]. See text for more detail.

Dataset	Brief Description	Pixel Acc.	Obj. Acc.
ssTEM	appearance	45.06%	10.01%
	shape	51.14%	15.37%
	context	72.03%	56.25%
	three features	75.07%	64.71%
hand	appearance	78.18%	82.07%
	shape: only single-scale	51.16%	73.33%
	shape	68.02%	75.69%
	context	74.46%	83.06%
	three features	79.05%	85.10%
	dedicated unsupervised seg.	56.99%	50.85%
serous	appearance	83.91%	81.43%
	shape	71.61%	64.07%
	context	65.15%	47.60%
	three features	85.11%	83.98%
	dedicated unsupervised seg.	62.92%	38.28%

consider all the complementary features. Since our discriminative learning approach is able to perform an implicit feature selection, through learning it usually allocates higher weights to the "good" features. Interestingly, even when some features fail or perform less well when being used alone, a further gain can usually be obtained by retaining those features: Empirically this phenomenon is observed for all three datasets. Consider the ssTEM dataset for example, by employing the best (context) feature, a pixel- (object-) level performance of about 72% (56%) is attained; which is much better than considering the other two features where the corresponding results are merely no more than 51% (15%). However, when considering all three features jointly, the performance is further improved to 75% (65%). We think this might be attributed to the complementary nature of these features.

As a comparison method, a state-of-the-art *dedicated unsupervised Segmenter* [13] has been implemented. This method contains a few steps including noise removal, Gaussian smoothing, and thresholding based on the color histogram. After converted to binary images, distance transform is applied, and object centers are detected by seed finding. Then the generalized Voronoi diagram is applied to separate the touching objects. Unfortunately we can not produce a reasonable result for ssTEM dataset using the method of [13], despite significant effort in tuning the internal parameters. We speculate this is because the mitochondria objects are not sufficiently distinct in any of the color channels. Notice for each of the datasets, the internal parameters are *manually* adjusted to attain best performance. Nevertheless, as demonstrated in Table 1, this method performs considerably inferior to those from our approach.

We also compare our results to those of [8], which describes a supervised method combining pixel classifier and watershed, and has its results on the serous dataset: This

method reports an accuracy of 93.67% when a K-means RGB is used, and 96.47% if a Bayes RGB is deployed instead. Note [8] uses the traditional pixel-level accuracy, as the average percentage of pixels being correctly classified for both foreground and background classes. Meanwhile our approach on serous dataset achieves a better result of 98.12% under this evaluation criteria (85.11% under our evaluation criteria). Note that for serous dataset the overall class distributions are 7% for the nuclei pixels and 93% for the background pixels. As explained previously, these scores therefore tend to be saturated and more biased toward the background class.

4 Outlook

For future work we plan to conduct extensive evaluation of our approach on different microscopic datasets, and extend to work with 3D.

References

1. Bengtsson, E., Wahlby, C., Lindblad, J.: Bobust cell image segmentation methods. Pattern Recognition and Image Analysis 14(2), 157–167 (2004)
2. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. IEEE TPAMI 23(11), 1222–1239 (2001)
3. Cardona, A., Saalfeld, S., Preibisch, S., Schmid, B., Cheng, A., Pulokas, J., Tomancak, P., Hartenstein, V.: An integrated micro- and macroarchitectural analysis of the drosophila brain by computer-assisted serial section electron microscopy. PLoS Biol. 8 (2010)
4. Chen, S., Gordon, G., Murphy, R.: Graphical models for structured classification, with an application to interpreting images of protein subcellular location patterns. J. Mach. Learn. Res. 9, 651–682 (2008)
5. Everingham, M., Gool, L.V., Williams, C., Winn, J., Zisserman, A.: The Pascal Visual Object Classes (VOC) challenge. Int. J. of Computer Vision 88(2), 303–338 (2010)
6. Felzenszwalb, P., McAllester, D., Ramanan, D.: A discriminatively trained, multiscale, deformable part model. In: Int. Conf. Computer Vision and Pattern Recognition (2008)
7. Levinshtein, A., Stere, A., Kutulakos, K., Fleet, D., Dickinson, S., Siddiqi, K.: Turbopixels: Fast superpixels using geometric flows. IEEE TPAMI 31, 2290–2297 (2009)
8. Lezoray, O., Cardot, H.: Cooperation of color pixel classification schemes and color watershed: a study for microscopical images. IEEE TIP 11(7), 783–789 (2002)
9. Lindeberg, T.: Edge detection and ridge detection with automatic scale selection. IJCV 30, 117–154 (1998)
10. Marcuzzo, M., Quelhas, P., Campilho, A., Mendonca, A.M., Campilho, A.: Automated arabidopsis plant root cell segmentation based on svm classification and region merging. Comput. Biol. Med. 39, 785–793 (2009)
11. Sivic, J., Russell, B., Efros, A., Zisserman, A., Freeman, W.: Discovering object categories in image collection. In: ICCV (2005)
12. Tsochantaridis, I., Joachims, T., Hofmann, T., Altun, Y.: Large margin methods for structured and interdependent output variables. J. Mach. Learn. Res. 6, 1453–1484 (2005)
13. Yu, W., Lee, H., Hariharan, S., Bu, W., Ahmed, S.: Evolving generalized voronoi diagrams of active contours for accurate cellular image segmentation. Cytometry 77, 379–386 (2010)

Fast Globally Optimal Segmentation of Cells in Fluorescence Microscopy Images

Jan-Philip Bergeest and Karl Rohr

University of Heidelberg, BIOQUANT, IPMB, and DKFZ Heidelberg
Dept. Bioinformatics and Functional Genomics, Biomedical Computer Vision Group

Abstract. Accurate and efficient segmentation of cells in fluorescence microscopy images is of central importance for the quantification of protein expression in high-throughput screening applications. We propose a new approach for segmenting cell nuclei which is based on active contours and convex energy functionals. Compared to previous work, our approach determines the global solution. Thus, the approach does not suffer from local minima and the segmentation result does not depend on the initialization. We also suggest a numeric approach for efficiently computing the solution. The performance of our approach has been evaluated using fluorescence microscopy images of different cell types. We have also performed a quantitative comparison with previous segmentation approaches.

1 Introduction

Cell nucleus segmentation is one of the most important tasks in analyzing and quantifying fluorescence microscopy images. In particular, in high-throughput applications, semi-automatic and manual analysis are not feasible because of the enormous amount of image data. Automatic methods are needed which efficiently deal with different cell types and image artifacts such as intensity inhomogeneities.

In recent years, different approaches for the segmentation of cell nuclei in fluorescence microscopy images have been introduced. Often thresholding approaches are applied which, however, suffer from intensity inhomogeneities within nuclei and over a whole image. To separate clustered nuclei, watershed-based techniques are frequently used (e.g. [13,7]). A main class of cell segmentation approaches is based on deformable models, which allow incorporation of a priori knowledge and can capture a wide spectrum of different shapes. One can distinguish between parametric models and implicit models (e.g. [12,4,9,14,8,10,5]). Implicit representations based on level sets have gained increased interest since topological changes can be handled naturally. In most of these approaches region-based energy functionals are employed. However, a disadvantage of previous level set approaches is that the underlying energy functionals lead to non-convex minimization problems. Hence, the optimization function has local minima and the global solution is generally not found (using local optimization methods which are typically applied). In addition, the segmentation result depends on the initialization.

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 645–652, 2011.

In this contribution, we introduce a new approach for cell nucleus segmentation in fluorescence microscopy images. Compared to previous work, our approach is based on two energy functionals that lead to convex minimization problems for which global solutions are determined. Also, we take advantage of the combination of the region-based functional of Chan-Vese [2] and the Bayesian functional proposed in [11] which was used for cell segmentation in [5]. A convex formulation for the region-based functional in [2] was derived in [6] and we employ this scheme for splitting clustered cell nuclei. In addition, we use the Bayesian functional in [11] to cope with intensity inhomogeneities. We reformulate this functional leading to a convex optimization problem. For both functionals we propose an efficient minimization scheme based on the Split Bregman method. We have successfully applied our approach to two different sets of 2D fluorescence microscopy images comprising different cell types and we have compared the results with previous segmentation approaches.

2 Cell Nuclei Segmentation

Our approach for the segmentation of cell nuclei in fluorescence microscopy images is based on two energy functionals. In this section, we first describe the original non-convex formulation of the functionals. Then we introduce a reformulation as convex optimization problem and we suggest an efficient numerical technique for finding the solution.

2.1 Non-convex Energy Functionals

The first functional used in our approach is the region-based energy functional proposed by Chan-Vese [2]:

$$E_1(\Theta, \partial\Omega) = \lambda \left(\kappa_0 \int_{\Omega_0} (I(\mathbf{x}) - \mu_0)^2 d\mathbf{x} + \kappa_1 \int_{\Omega_1} (I(\mathbf{x}) - \mu_1)^2 d\mathbf{x} \right) + Per(\Omega_1), \tag{1}$$

with $\Theta = (\mu_0, \mu_1)$, where μ_0 and μ_1 are the mean intensities of the background region Ω_0 and the foreground region Ω_1, respectively. $\partial\Omega$ denotes the boundaries between the regions Ω_i, $I(\mathbf{x})$ are the image intensities at position $\mathbf{x}$, $Per(\Omega_1)$ is the perimeter of Ω_1, and $\kappa_0, \kappa_1, \lambda$ are weighting factors. As the second functional, we use the one proposed in [11], which was derived based on a Bayesian approach:

$$E_2(\Theta, \partial\Omega) = \lambda \left(\int_{\Omega_0} -\log P(I(\mathbf{x})|\Omega_0) d\mathbf{x} + \int_{\Omega_1} -\log P(I(\mathbf{x})|\Omega_1) d\mathbf{x} \right) + Per(\Omega_1), \tag{2}$$

with $\Theta = (\mu_0, \mu_1, \sigma_0, \sigma_1)$, where μ_i and σ_i are the mean intensities and standard deviations of the regions Ω_i. $P(I(\mathbf{x})|\Omega_i)$ is the conditional probability that pixel $\mathbf{x}$ with intensity $I(\mathbf{x})$ belongs to region Ω_i. Here, we assume a Gaussian distribution $P(I(\mathbf{x})|\Omega_i) = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(I(\mathbf{x})-\mu_i)^2}{2\sigma_i^2}}$. An advantage of E_2 compared to E_1 is that the weights κ_0 and κ_1 are not needed, but are implicitly included via

the standard deviations σ_i, which are estimated from the image data. Using a level set representation for E_1 and E_2 and applying the Euler-Lagrange equation leads to the gradient flow for the level set function ϕ

$$\frac{\partial \phi(\mathbf{x})}{\partial t} = \left(-\lambda r_j + \nabla \cdot \frac{\nabla \phi}{|\nabla \phi|} \right) H'(\phi(\mathbf{x})), \tag{3}$$

with $H(x)'$ being the derivative of the Heaviside function $H(x) = \{0 \text{ if } x < 0,\ 1 \text{ if } x \geq 0\}$. The term r_j corresponds to the external image forces of the two energies $E_j, j = 1, 2$, which are defined as

$$r_1 = \kappa_1 (I(\mathbf{x}) - \mu_1)^2 - \kappa_0 (I(\mathbf{x}) - \mu_0)^2 \tag{4}$$

$$r_2 = \log P(I(\mathbf{x})|\Omega_1) - \log P(I(\mathbf{x})|\Omega_0) \tag{5}$$

The parameter vector Θ can be computed directly:

$$\mu_0 = \frac{\int I(\mathbf{x})(1 - H(\phi(\mathbf{x})))d\mathbf{x}}{\int (1 - H(\phi(\mathbf{x})))d\mathbf{x}} \qquad \mu_1 = \frac{\int I(\mathbf{x}) H(\phi(\mathbf{x}))d\mathbf{x}}{\int H(\phi(\mathbf{x}))d\mathbf{x}}$$

$$\sigma_0^2 = \frac{\int (I(\mathbf{x}) - \mu_0)^2 (1 - H(\phi(\mathbf{x})))d\mathbf{x}}{\int (1 - H(\phi(\mathbf{x})))d\mathbf{x}} \qquad \sigma_1^2 = \frac{\int (I(\mathbf{x}) - \mu_1)^2 H(\phi(\mathbf{x}))d\mathbf{x}}{\int H(\phi(\mathbf{x}))d\mathbf{x}}$$

2.2 Convex Energy Functionals

In [1], it has been shown that certain non-convex minimization problems can be reformulated as convex problems. There, a convex formulation for E_1 in (1) was derived. In our approach, we use this formulation and we also derive a convex formulation for E_2 in (2). Following [1], the Heaviside function in (3) is omitted since (3) and the following gradient descent equation have the same steady state solution

$$\frac{\partial \phi(\mathbf{x})}{\partial t} = -\lambda r_j + \nabla \cdot \frac{\nabla \phi}{|\nabla \phi|}. \tag{6}$$

Then, the corresponding energy functional can be stated as

$$E_j(\Theta, \phi) = \lambda \langle \phi, r_j \rangle + |\nabla \phi|_1, \tag{7}$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product, $|\cdot|_1$ is the $L1$-norm, and $|\nabla \phi|_1$ corresponds to $Per(\Omega_1)$. By restricting ϕ to lie in a finite interval, e.g. [0, 1], the global minimum can be guaranteed by solving the following convex problem for the normalized ϕ_n:

$$\min_{0 \leq \phi_n \leq 1} E_j(\Theta, \phi_n) = \lambda \langle \phi_n, r_j \rangle + |\nabla \phi_n|_1 \tag{8}$$

2.3 Split Bregman Method

To solve (8) we use the Split Bregman method. This method is a general technique for efficiently solving $L1$-regularized problems and for iteratively finding extrema of convex functionals [6]. The method consists of variable splitting

and Bregman iteration. Variable splitting is achieved by introducing the auxiliary vector $\boldsymbol{d}$ (with dimension according to the image domain) and by using a quadratic penalty term to enforce the constraint $\boldsymbol{d} = \nabla\phi_n$. For (8) this leads to

$$(\phi_n^*, \boldsymbol{d}^*) = \arg \min_{0 \leq \phi_n \leq 1, \boldsymbol{d}} \left(\lambda \langle \phi_n, r_j \rangle + |\boldsymbol{d}|_1 + \frac{\nu}{2} |\boldsymbol{d} - \nabla\phi_n|_2^2 \right), \tag{9}$$

where ϕ_n^* and $\boldsymbol{d}^*$ denote the iteratively computed solution. In (9) $\nabla\phi_n$ is no longer associated with the $L1$-norm, however, the constraint $\boldsymbol{d} = \nabla\phi_n$ is only weakly enforced. To enforce the constraint exactly, the Bregman iteration technique is applied. With this technique a vector $\boldsymbol{b}$ is included in the penalty function and an alternating minimization is carried out:

$$(\phi_n^k, \boldsymbol{d}^k) = \arg \min_{0 \leq \phi_n \leq 1, \boldsymbol{d}} \left(\lambda \langle \phi_n, r_j \rangle + |\boldsymbol{d}|_1 + \frac{\nu}{2} |\boldsymbol{d} - \nabla\phi_n - \boldsymbol{b}^{k-1}|_2^2 \right), \tag{10}$$

$$\boldsymbol{b}^k = \boldsymbol{b}^{k-1} + \nabla\phi_n^k - \boldsymbol{d}^k, \tag{11}$$

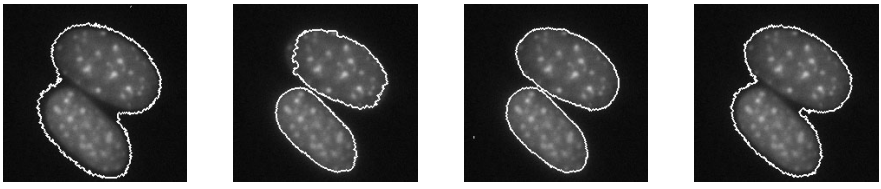
where ϕ_n^k and $\boldsymbol{d}^k$ represent the solution at iteration k. The problem in (10),(11) is first solved w.r.t. ϕ_n, while $\boldsymbol{d}$ and $\boldsymbol{b}$ are fixed and second w.r.t. $\boldsymbol{d}$. In our case, we use a fast iterative Gauss-Seidel solver for the first minimization. The second minimization is obtained explicitly using the vector-valued shrinkage operator:

$$\boldsymbol{d}^k = \max\{|\boldsymbol{b}^{k-1} + \nabla\phi_n^k|_2 - \nu, 0\} \frac{\boldsymbol{b}^{k-1} + \nabla\phi_n^k}{|\boldsymbol{b}^{k-1} + \nabla\phi_n^k|_2} \tag{12}$$

2.4 Globally Optimal Cell Segmentation

Our approach for cell nuclei segmentation combines the two convex energy functionals E_1 and E_2 in (8). The functional E_2 is used to deal with intensity inhomogeneities over the whole image and to segment cell nuclei with varying intensities. E_1 is used to split cell nuclei which have been falsely merged using E_2 and is minimized independently for all regions that were segmented using E_2. Note, that doing this we use multiple level sets for the segmentation of an image. This is similar to multiple level set approaches (e.g. [12,4,9,14,8]). However, in our case we do not need a coupling term to prevent level sets from merging because we perform the minimization within previously segmented regions.

Our approach consists of three steps. In the first step, we perform a segmentation of the whole image using the functional E_2. Since this functional comprises parameters for both the mean intensities and the variances of objects we can deal with varying background intensities and with cell nuclei with varying intensities. However, cell nuclei that are close to each other are likely to be merged in this step (see Fig. 1a). In the second step, we minimize E_1 for each of the segmented cell nuclei regions in the first step. This allows splitting of clustered cell nuclei (see Fig. 1b). In the third step, we use E_2 again but this time for the segmented objects from the second step within a region of interest around the object. The third step is needed because the second step cannot deal with inhomogeneities within cell nuclei. The result of the third step in Fig. 1c shows that the segmentation at the border of the nuclei is more accurate and smoother compared to



(a) (b) (c) (d)

Fig. 1. Original images overlaid with contours of the segmentation results. Result after the first step (a), the second step (b), and the third step (c). (d) shows the result when using E_2 in the second step.

the result after the second step. Note, that if we would use E_2 in the second step then clustered nuclei could not be separated because the contrast in the region between the cells is not large enough (see Fig. 1d).

3 Experimental Results

We have applied our approach to two sets of 2D fluorescence microscopy images of cell nuclei of different cell types from [3] for which ground truth is available. The first data set consists of 48 images with a size of 1349×1030 pixels which include in total 1831 U20S Hoechst stained cell nuclei (see Fig. 2a). The second data set contains 49 images with a size of 1344×1024 pixels comprising in total 2178 NIH3T3 Hoechst stained nuclei (see Fig. 2d). Note, that several images in the second data set are heavily affected by intensity inhomogeneities and visible artifacts. Therefore, automatic analysis of the second set is more challenging compared to the images in the first set.

To evaluate the performance of our approach we determined region-based and contour-based measures. As region-based measure we used the Dice coefficient and as contour-based measures we employed the normalized sum of distances (NSD) [3] and the Hausdorff distance. The Dice coefficient is defined as $Dice(R,S) = \frac{2|R \cap S|}{|R|+|S|}$, where R is the binary reference image, and S is the binary segmented image. The NSD is defined as $\mathrm{NSD}(R,S) = \frac{\sum_{i \in R \cup S \setminus R \cap S} D(i)}{\sum_{i \in R \cup S} D(i)}$, where $D(i)$ is the minimal Euclidean distance of pixel i to the contour of the reference object. The Hausdorff distance is defined as $h(R,S) = \max_{i \in S_c}\{D(i)\}$, with S_c being the contour of the segmented object. We also used two detection measures, namely the number of false positives (FP) and the number of false negatives (FN). FP corresponds to spuriously segmented nuclei and FN corresponds to nuclei that have not been segmented. In all our experiments we used $\nu = 10$ and $\kappa_0 = 1, \kappa_1 = 1$ in (10),(11). In the first and second step of our approach we used $\lambda = 10000$ and in the third step we chose $\lambda = 1000$.

Table 1 shows the results of our approach (after 1 step and after all 3 steps) for the different performance measures averaged over all images of each data set. As a comparison we also show the results for Otsu thresholding as well as the

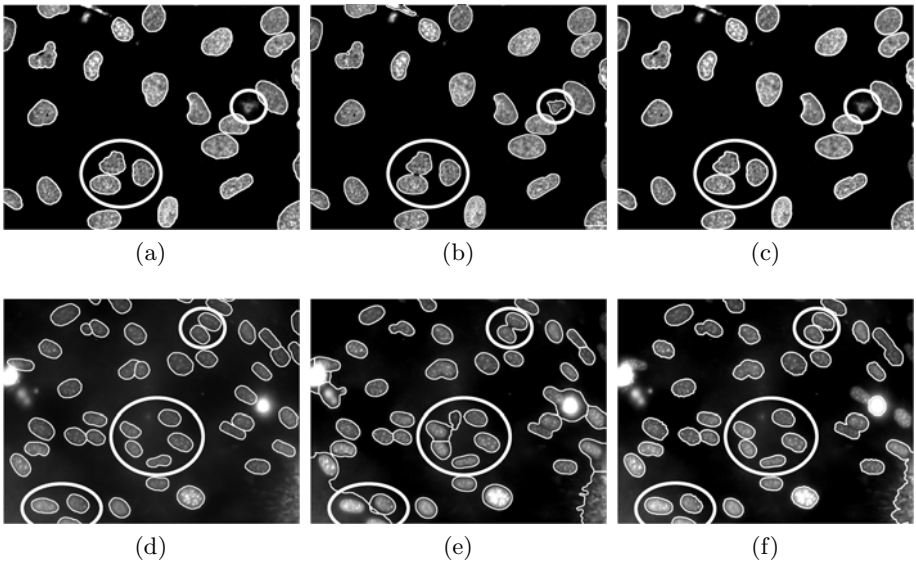


Fig. 2. Original images with overlaid contours of segmentation results. First row: U20S cells, second row: NIH3T3 cells; first column: ground truth, second column: merging algorithm [7], third column: our approach.

watershed algorithm and the region merging algorithm [7] applied to the mean thresholded image. The latter approach uses an improved watershed algorithm followed by statistical model-based region merging to cope with oversegmentation and yielded the best results in the study in [3]. Also, we included the result of using the non-convex functional E_2 in (2). In addition, we included the result of manual segmentation (for a subset of 5 images for each data set) by a different observer than the one who provided the ground truth (second row, Manual). It can be seen that our approach yields the best result for the Dice coefficient and the NSD for both data sets, and that the result is comparable to the result by the different observer (Manual). In particular, for the more challenging NIH3T3 images we obtain significantly better results than previous approaches. Regarding the Hausdorff distance, our approach yields better results for the NIH3T3 cell images, while the results are equally good as those for the algorithm in [7] for the U20S images. Furthermore, it can be seen that our approach (3 steps) yields better results than our 1 step approach and thus the multiple step scheme is advantageous. However, the 1 step approach still yields better results than using the non-convex functional E_2 in (2). For the FP value the result is equally good or similar to the best results of the other approaches. Regarding the FN value we obtain for the U20S cell images similar results as for the other approaches, while for the NIH3T3 cell images we obtain intermediate results. However, note that in biological applications a small FP value is more important than a small FN value. The reason is that it is often necessary to quantify the signal intensity in a second channel within segmented nucleus regions, and FP regions

Table 1. Quantitative results for the different approaches

Approach	U20S cells (48 images)					NIH3T3 cells (49 images)				
	Dice	NSD	Hausdorff	FP	FN	Dice	NSD	Hausdorff	FP	FN
Manual	0.93	0.04	9.8	0.6	2.2	0.87	0.07	12.1	0.0	3.2
Otsu	0.87	0.12	34.8	0.3	5.5	0.64	0.35	36.7	1.7	26.4
Watershed	0.69	0.36	34.3	1.9	3.0	0.62	0.37	19.1	11.6	5.5
Merging algorithm [7]	0.92	0.08	13.3	1.0	3.3	0.70	0.28	19.0	7.0	5.8
Non-convex E_2 in (2)	0.88	0.15	25.2	1.5	2.8	0.76	0.24	21.9	3.5	5.2
Our approach (1 step)	0.93	0.11	18.1	1.4	3.2	0.84	0.23	21.0	3.3	8.6
Our approach (3 steps)	0.94	0.06	13.3	0.5	3.9	0.83	0.14	16.5	1.7	11.3

have a strong influence on the statistical analysis. Compared to the approach in [7] our approach allows better splitting of clustered cell nuclei and spurious objects are not included (see Fig. 2b,c), and our approach yields a more accurate segmentation of cell nuclei shapes (see Fig. 2e,f).

In our approach, the minimization of E_2 in (8) on the whole image converges after about 5 iterations and the computation time is approximately 15 seconds per image. In comparison, using a standard level set scheme for the non-convex energy E_2 in (2) with gradient descent optimization needs about 400 iterations and the computation time is about 9 minutes. Thus, our approach is significantly faster. As convergence criterion we used the average Euclidean distance between the contours of segmented objects in two successive iterations.

4 Discussion

We have introduced a new approach based on active contours for cell nuclei segmentation in fluorescence microscopy images. Our approach employs two convex energy functionals for which globally optimal solutions are determined. By combining the two functionals we can cope with clustered cells, different cell types, and intensity inhomogeneities. To minimize the energy functionals we suggested using the Split Bregman method which significantly reduces the computation time compared to standard level set approaches. Thus, our approach is suitable for high-throughput applications and very large data sets. We have demonstrated the applicability of our approach using 97 real fluorescence microscopy images comprising in total 4009 cells of two different cell types. It turned out that our approach yields superior results compared to previous approaches.

Acknowledgments. We thank Luis Pedro Coelho (CMU) for help with his software. Support of the BMBF project FANCI (SysTec) is gratefully acknowledged.

References

1. Chan, T.F., Esedoglu, S., Nikolova, M.: Algorithms for finding global minimizers of image segmentation and denoising models. SIAM J. Appl. Math. 66, 1632–1648 (2006)
2. Chan, T.F., Vese, L.A.: Active contours without edges. IEEE Trans. Image Process 10(2), 266–277 (2001)
3. Coelho, L.P., Shariff, A., Murphy, R.F.: Nuclear segmentation in microscope cell images: a hand-segmented dataset and comparison of algorithms. In: ISBI 2009, pp. 518–521 (2009)
4. Dufour, A., Shinin, V., Tajbakhsh, S., Guillen-Aghion, N., Olivo-Marin, J.C., Zimmer, C.: Segmenting and Tracking Fluorescent Cells in Dynamic 3-D Microscopy With Coupled Active Surfaces. IEEE Trans. Image Process 14(9), 1396–1410 (2005)
5. Dzyubachyk, O., van Cappellen, W.A., Essers, J., Niessen, W.J., Meijering, E.H.W.: Advanced Level-Set-Based Cell Tracking in Time-Lapse Fluorescence Microscopy. IEEE Trans. Med. Imag. 29(3), 852–867 (2010)
6. Goldstein, T., Bresson, X., Osher, S.: Geometric Applications of the Split Bregman Method: Segmentation and Surface Reconstruction. J. of Scientific Computing 45(1), 272–293 (2010)
7. Lin, G., Adiga, U., Olson, K., Guzowski, J.F., Barnes, C.A., Roysam, B.: A hybrid 3D watershed algorithm incorporating gradient cues and object models for automatic segmentation of nuclei in confocal image stacks. Cytometry A 56(1), 23–36 (2003)
8. Mosaliganti, K., Gelas, A., Gouaillard, A., Noche, R., Obholzer, N., Megason, S.: Detection of Spatially Correlated Objects in $3D$ Images Using Appearance Models and Coupled Active Contours. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5762, pp. 641–648. Springer, Heidelberg (2009)
9. Nath, S.K., Palaniappan, K., Bunyak, F.: Cell segmentation using coupled level sets and graph-vertex coloring. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4190, pp. 101–108. Springer, Heidelberg (2006)
10. Padfield, D., Rittscher, J., Thomas, N., Roysam, B.: Spatio-temporal cell cycle phase analysis using level sets and fast marching methods. Med. Image Anal. 13(1), 143–155 (2009)
11. Rousson, M., Deriche, R.: A Variational Framework for Active and Adaptative Segmentation of Vector Valued Images. In: Proc. Workshop Motion Video Comput., pp. 56–62. IEEE Computer Soc., Los Alamitos (2002)
12. Ortiz de Solorzano, C., Malladi, R., Lelievre, S., Lockett, S.J.: Segmentation of nuclei and cells using membrane related protein markers. J. Micr. 201(3), 404–415 (2001)
13. Wählby, C., Lindblad, J., Vondrus, M., Bengtsson, E., Björkesten, L.: Algorithms for cytoplasm segmentation of fluorescence labelled cells. Analytical Cellular Pathology 24(2), 101–111 (2002)
14. Yan, P., Zhou, X., Shah, M., Wong, S.T.C.: Automatic segmentation of high-throughput RNAi fluorescent cellular images. IEEE Trans. Inf. Techn. in Biomed 12(1), 109–117 (2008)

Carving: Scalable Interactive Segmentation of Neural Volume Electron Microscopy Images

C.N. Straehle[1], U. Köthe[1], G. Knott[2], and F.A. Hamprecht[1]

[1] University of Heidelberg, Heidelberg, Germany
HCI, Speyerer Strasse 6, D-69115 Heidelberg
fred.hamprecht@iwr.uni-heidelberg.de
[2] Ecole Polytechnique Fédérale, Lausanne, Switzerland

Abstract. Interactive segmentation algorithms should respond within seconds and require minimal user guidance. This is a challenge on 3D neural electron microscopy images. We propose a supervoxel-based energy function with a novel background prior that achieves these goals. This is verified by extensive experiments with a robot mimicking human interactions. A graphical user interface offering access to an open source implementation of these algorithms is made available.

Keywords: electron microscopy, seeded segmentation, interactive segmentation, graph cut, watershed, supervoxel.

1 Introduction

Electron microscopy has provided images with revealing resolution in *two* dimensions since the mid-20th century. More recent volume imaging techniques such as Focused Ion Beam Scanning Electron Microscopy (FIBSEM), however, yield images with an isotropic resolution of a few nm in all *three* dimensions [11], see Fig. 1. Full automation is required for the analysis of very large scale experiments, e.g. in connectivity studies. Several recent methods [1,16,15,10,6,9] already take advantage of the isotropy in new datasets. However, the error rates of these methods still fall short of human performance. In contrast, interactive methods allow live user corrections until the desired quality is achieved. This is useful during rapid exploration of new data, but even more important for generating the ground truth needed for training and validation of automated algorithms.

Interactive or "seeded" segmentation methods come in two main flavors: deformable models ("snakes" e.g. [7,8,17]) and random fields (e.g. [18,19]). The former are particularly useful when objects are characterized by recurring shape properties that constrain the segmentation. Slice-oriented versions of this approach (using 3D tracking of 2D models) were successfully used on anisotropic neuron data [7], but truly 3-dimensional modeling is challenged by the intricate geometry of neurons, which bend and branch out in complex ways. Random fields naturally lend themselves to 3D modeling and thus seem a natural choice

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 653–660, 2011.

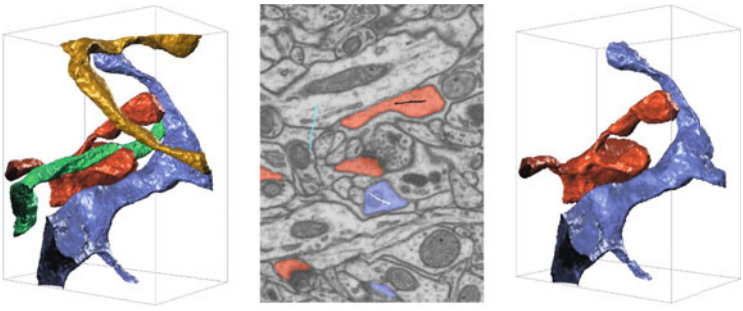


Fig. 1. (left) Small subset of the ground truth segmentation. (center) Slice from FIBSEM raw data, with seeds provided by the user (background: cyan, foreground: black,white) and the resulting segmentation (orange, blue). (right) 3D rendering of the segmentation shown in (b). (Rendering done with V3D [13]).

for isotropic neuron data. However, different neurons of the same type are indistinguishable by their local appearance (intensity and texture), so that local class probabilities are hard to define. This also causes strong shrinking bias [18,19]. Our main contributions and insights are a graph cut formulation and a watershed with:

- a problem-specific objective function whose potentials are biased toward the background region, so that uncertain voxels are preferably assigned to the background. We show that this simple idea works surprisingly well on neural data.
- a *supervoxel* approximation [14] of the original problem defined on a voxel grid. We achieve 100-fold speed-ups with virtually identical results.
- an edge weight parametrization which overcomes the shrinking bias. This is complementary to, and arguably simpler than, gradient fluxes [18,19].
- an extensive quantitative comparison of graph cut and watershed results with ground truth on 20 dendrites by means of a *robot user* that mimics human seed selection.

Satisfactory segmentations of individual 3D objects from a 500^3 FIBSEM volume can typically be achieved with 2-4 seeds with a response time of ca. 1 second per click (after $\approx$ 3 minutes of preprocessing at program startup).

2 The Optimization Problem: Basic Definition, Enhancements and Efficient Solution via Supervoxels

Reference [3] is the culmination of a series of important papers relating graph cut, random walk, shortest paths and watershed in a unified seeded segmentation framework, the "Power watershed". We will start from these insights, and represent an image as a weighted graph $G = (V, E)$ of pixels $v \in V$ and their neighborhood relations $e \in E$. A segmentation is represented by a set of labels $\mathbf{x}$

associated with the vertices V. A label x_i is coupled to the underlying observations at node v_i by a node weight w_i, and to its neighbors by weights $w_{ij} \in [0,1]$ assigned to the edges e_{ij}. Large edge weights express a pronounced penchant for the incident vertices to share the same label. Finally, user input is added in the form of seeds $y_i \in \{0,\ldots,|C|-1\}$ in an interactive fashion, where C is the set of distinct regions. Graph cut can handle only $|C| = 2$ distinct regions (foreground vs. background), while the watershed can account for an arbitrary number of regions. A segmentation is then given by the minimizer

$$\operatorname{argmin}_{\mathbf{x}} \sum_{e_{ij} \in E} w_{ij}^p |x_i - x_j|^q + \sum_{v_i \in V} w_i^p |x_i - y_i|^q \tag{1}$$

with graph cut ($p = q = 1, y_i, x_i \in \{0,1\}$), random walk ($p = 1, q = 2$, $y_i \in \{0,1\}$, $x_i \in [0,1]$) and watershed ($p \to \infty, q = 1, y_i, x_i \in \{0,\ldots,|C|-1\}$) emerging as special cases.

2.1 The Crucial Choice of Weight Functions

Given the above framework, the developer has to choose p, q and, importantly, the weights w. Node weights are often used to couple the node label to the appearance of the underlying region, as in GrabCut. This is not viable in the data studied here: for instance, the texture, color, etc. of one dendrite are indistinguishable from that of others in the vicinity. The modeling effort must hence concentrate on region boundaries. This can happen indirectly, through nonlocal geometric terms which can be cleverly encoded in node potentials [12,18], or directly through boundary detectors. These are often implemented through $w_{ij} = \exp\left(-\beta |\nabla I|^2_{ij}\right)$ where $|\nabla I|^2_{ij}$ is the squared image gradient between nodes i and j. As noted in [3], the parameter β plays a crucial role. It not only indicates what gradients are deemed significant, but is directly related to the parameter p from (1): increasing β is tantamount to increasing p!

Since the staining in the neural EM images studied here can directly be interpreted as an edge map, we instead define

$$w_{ij} = \exp\left(-\beta \left(I_i + I_j\right)/2\right) \tag{2}$$

where I is an estimate for membrane probability. Besides the (inverted) EM image itself, $I \in [0,1]$ can be any membrane indicator, for instance a discriminative classifier [1,15] or the first eigenvalue of the Hessian matrix of the membrane image. We tried both, but concentrate on the latter here because it needs no training and allows us to focus on the improvements described below.

The Impact of β. Changing β in (2) has important implications for graph cut: When using $\beta = 1$ and unary weights $w_i = 0$ (except at locations with user seeds), severe shrinking bias [12] occurs: the cheapest segmentation boundary tightly encloses either the foreground or background seeds, assigning nearly the complete volume to the other class. Increasing β to around 100 (and thus raising p 100-fold) implies that erroneous cuts within regions of low membrane strength

become very expensive, while label transitions on membranes remain cheap. In our data, this is sufficient to obviate more complex modeling, e.g. in terms of gradient flux unary potentials [18].

In contrast, changing β in the watershed amounts to a monotonous height transformation of the landscape that this algorithm implicitly sees, without however changing its salient features: the locations and ordering of the minima, maxima, saddle points and, most importantly, the watersheds do not change. The same conclusion – invariance of the watershed algorithm to the choice of β – can be reached by a study of the limit $p \to \infty$ in (2).

Background Bias. Assume that an appropriate value of β has been found. If a single foreground and a single background seed of similar size are given, one would expect around half of all neurons to be assigned to either class. This is borne out in practice. To achieve something closer to the desired result, namely a segmentation of one neuron versus all others, we propose to use a small bias favoring the background. For the graph cut, this bias is easily incorporated by assuming an implicit background label $y_i = 0$ for all unseeded points, and adding a small unary weight $w_i = \alpha$ to these points. The quality of the resulting segmentation is robust w.r.t. α across three orders of magnitude: choosing $\alpha \in [10^{-5}, 10^{-3}]$ results in much better segmentations than $\alpha = 0$ (results not shown).

The above recipe does not work for the watershed. Instead, we build the bias directly into the priority queue that controls the region assignment order of the algorithm: In every iteration, there is a set of regions (initially the seeds), and all edges connecting these regions to unlabeled nodes are ordered by weight w_{ij}. The end node of the most expensive edge is then assigned to the region where the edge starts. In this way, watersheds cut cheap edges, minimizing (1). Clearly, manipulations of the unary weights w_i and monotonous transformations of the binary weights w_{ij} have no influence on the outcome. Therefore, we change (2) so that assignments to the background class are preferred: $w_{ij}(x_i) = \exp\left(-\gamma(x_i)\,\beta\,(I_i + I_j)\,/2\right)$ with $0 < \gamma(x_i = 0) < \gamma(x_i \neq 0) \leq 1$. Qualitatively speaking, the dams in the flooding metaphor of the watershed algorithm appear lower to the background class.

2.2 Speedup by Coarse-Graining: Supervoxels

All algorithms studied here are too slow, in their native implementations, for a truly interactive experience when working with 3D volumes of the order of 512^3 voxels on a standard desktop PC. We hence suggest a coarse graining which is a heuristic approximation for graph cut and random walker, but does *not* affect the solution of the watershed. The simplification is best explained for random walk segmentation, but applies to the other cases as well: the labels computed by the random walker change little (cf. [4]) over homogeneous regions (where the binary weights are large), and abruptly near boundaries (where the binary weights are small). One natural simplification of the problem is to reduce the number of unknowns in the linear system of equations by grouping such pixels as are expected to have very similar labels in the solution. Conceptually, this amounts to setting some of the binary weights to very large values, and hence constraining

all pixels within a group or "superpixel" to have the same label. More explicitly, we propose to augment (1) with a set of constraints $x_i = x_j | i, j \in s_k$ for a suitable partitioning $\bigcup_k s_k = V$, $s_k \cap s_l = \emptyset$ of the original graph G.

The quality of this approximation crucially depends on the partitioning used, and finding a good partitioning looks like a daunting task: it should on the one hand be cheap to compute, and on the other hand anticipate the solution of the segmentation algorithm even before that is executed. However, based on the analysis of the powerwatershed energy (1), we argue that the catchment basins of the watershed are a useful proxy: remember that both graph cut and random walk require very pronounced edge weights (large β, or large p) to yield sensible solutions on the type of data used here. Such large powers $p \approx 100$ are an approximation to the special case $p \to \infty$ which is solved by the watershed. Conversely, the catchment basins $\{s_k\}$ of an (unseeded) watershed based on the same edge weights are regions in which the labels of graph cut or random walk are relatively homogeneous.

Making this simplifying assumption, we obtain a supervoxel graph $G'(E', V')$ in which the vertices $s_k \in V'$ represent the individual catchment basins of the original graph G, and the weight $w'_{k,l}$ of an edge in E' is given by the sum over the edge weights connecting the two catchment basins s_k, s_l from the original graph. This approximation makes for large savings, especially for the segmentation of multiple objects in the same data set for which other strategies such as reusing the residual flow of a graph cut computation cannot be employed.

3 Experiments

Data. The $600 \times 800 \times 409$ dataset shows neural tissue imaged with a FIBSEM instrument [11]. To evaluate the algorithms objectively, we designed an **Interactive Segmentation Robot.** [5]. The automaton emulates the human seeding strategy for different parameters and objects. Given ground truth (Fig. 1), the robot seeks to interactively segment a single object of interest using the following strategy:

1. Calculate the set differences between ground truth and current segmentation. Place a correcting single voxel seed in the center (maximum of the Euclidean distance transform) of the largest false region.
2. Re-run the segmentation algorithm with the new set of seeds.
3. Iterate until convergence to ground truth.

The convergence criterion is that the detected boundaries are within three pixels of the true boundaries (which roughly corresponds to the accuracy of the ground truth). Note that the estimated number of seeds required to reach a good segmentation is conservative because the robot only labels a single voxel at a time, while most humans would provide extended brush strokes which are potentially more informative.

All tests were executed on an Intel iCore 7 machine with 2.4GHz. Unfortunately, the random walker was too slow for an interactive procedure even in the

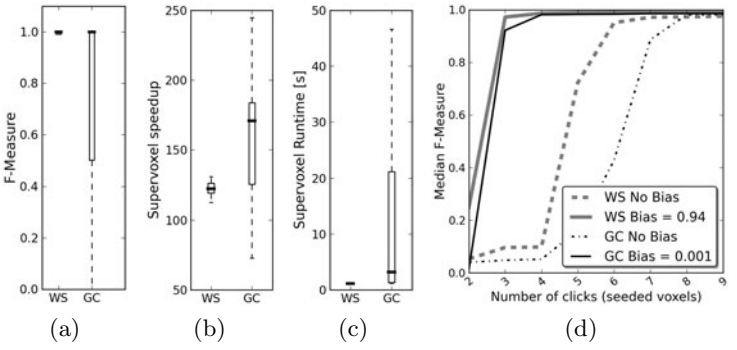


Fig. 2. (a) Agreement between voxel and supervoxel variants of the algorithms. (b) Speedup factor due to supervoxels. (c) Distribution of absolute response times to each click (after one-time preprocessing), on a 600×800×409 volume. (d) Agreement between ground truth and robot segmentations, as a function of the number of interactions.

supervoxel implementation, so the corresponding results are omitted. The graph cut problem was solved using the public code of [2]. The free parameters of the algorithms (scale of the Hessian matrix feature, exponent p in (1), strength of background bias $\gamma(x_i)$, α) were optimized on a volume that does not overlap the test region. Interestingly, the best choice for p was $p \approx 100$, confirming our discussion about the impact of this parameter in Section 2.

Results. The first experiment (Fig. 2a) investigates whether supervoxel segmentation is a valid approximation of voxel-level segmentation. This can be affirmed for the watershed (the observed minor differences can be traced to different implementations of the priority queue). For the graph cut, the results are a little surprising: while there is a discrepancy between voxel and supervoxel implementations, it is actually the latter that performs better! The reason is the shrinking bias: since the robot labels individual voxels, graph cut sometimes ignores these seeds by turning them into tiny "islands". This phenomenon does not occur when an entire supervoxel is seeded. The second experiment compares the speed of the supervoxel algorithms relative to their voxel-level implementations (Fig. 2b) and the absolute run time (Fig. 2c). We observe a median speed-up of around 120 for the watershed and around 170 for graph cut. Once the supervoxel graph has been constructed – this is a matter of minutes – the turnaround time for a user interaction is of the order of (sometimes many) seconds for the graph cut, and consistently around one second for the watershed. The third experiment (Fig. 2d) shows the number of clicks (single voxel seeds) given by the robot following the segmentation strategy and the ground truth agreement of the resulting segmentations. The figure shows the median over the 20 largest objects in a 300^3 subregion of the volume. Using a background bias clearly reduces the number of interactions required. For a qualitative impression, exemplary carving results are shown in Figs. 1 and 3.

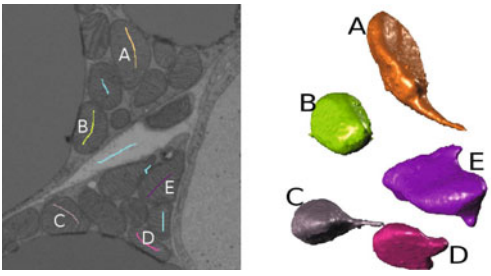


Fig. 3. The proposed approach works well on objects other than dendrites, here: mitochondria. Shown are (left): slice view of the raw data and user given segmentation seeds (background seed: cyan) and (right): the resulting segmentation for the foreground seeds.

4 Conclusions

We have systematically evaluated two well known algorithms (watersheds and graph cut) with regard to their applicability to interactive neuron segmentation in FIBSEM volume data. Our experiments have shown that the supervoxel approximation can be safely used as an approximation of the voxel-level energy function. This is in part due to the choice of a very large power $p \approx 100$ in (1). The main effect of this choice is to counter the shrinking bias of the graph cut, but it also serves to minimize the error arising from grouping voxels into supervoxels. The simple idea of preferring background over foreground assignments effectively helps ensuring that each foreground region consists of a single neuron only – the number of required user interactions is reduced considerably relative to unbiased assignments, resulting in correct segmentations after only a handful of mouse clicks. Moreover, no additional calculations are required to realize this behavior. All results were confirmed by extensive experiments using an objective segmentation robot. It applies a transparent seed placement strategy that emulates the actions of a human user.

Minimizing user effort and algorithm response times allows to analyze large data sets interactively, i.e. with immediate feedback on segmentation quality. The described supervoxel algorithms can be downloaded, along with a graphical user interface, as an open source program from http://www.ilastik.org/carving.

References

1. Andres, B., Köthe, U., Helmstaedter, M., Denk, W., Hamprecht, F.A.: Segmentation of SBFSEM volume data of neural tissue by hierarchical classification. In: Rigoll, G. (ed.) DAGM 2008. LNCS, vol. 5096, pp. 142–152. Springer, Heidelberg (2008)
2. Boykov, Y., Kolmogorov, V.: An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. IEEE TPAMI 26, 1124–1137 (2004)
3. Couprie, C., Grady, L., Najman, L., Talbot, H.: Power watershed: A unifying graph-based optimization framework. IEEE TPAMI 33, 1384–1399 (2011)

4. Grady, L.: Random walks for image segmentation. IEEE TPAMI 28, 1768–1783 (2006)
5. Gulshan, V., Rother, C., Criminisi, A., Blake, A., Zisserman, A.: Geodesic star convexity for interactive image segmentation. In: IEEE CVPR 2010, pp. 3129–3136 (2010)
6. Jain, V., Bollmann, B., Richardson, M., Berger, D., Helmstaedter, M., Briggman, K., Denk, W., Bowden, J., Mendenhall, J., Abraham, W., Harris, K., Kasthuri, N., Hayworth, K., Schalek, R., Tapia, J., Lichtman, J., Seung, H.: Boundary learning by optimization with topological constraints. In: IEEE CVPR 2010, pp. 2488–2495 (2010)
7. Jeong, W., Beyer, J., Hadwiger, M., Blue, R., Law, C., Vazquez, A., Reid, C., Lichtman, J., Pfister, H.: Ssecrett and neurotrace: Interactive visualization and analysis tools for large-scale neuroscience datasets. IEEE Comput Graph 30, 58–70 (2010)
8. Jurrus, E., Hardy, M., Tasdizen, T., Fletcher, P., Koshevoy, P., Chien, C., Denk, W., Whitaker, R.: Axon tracking in serial block-face scanning electron microscopy. Med. Image Anal. 13(1), 180–188 (2009)
9. Jurrus, E., Paiva, A., Watanabe, S., Anderson, J., Jones, B., Whitaker, R., Jorgensen, E., Marc, R., Tasdizen, T.: Detection of neuron membranes in electron microscopy images using a serial neural network architecture. Med. Image Anal. 14(6), 770–783 (2010)
10. Kaynig, V., Fuchs, T., Buhmann, J.: Neuron geometry extraction by perceptual grouping in sstem images. In: IEEE CVPR 2010, pp. 2902–2909 (2010)
11. Knott, G., Marchman, H., Wall, D., Lich, B.: Serial section scanning electron microscopy of adult brain tissue using focused ion beam milling. J. Neurosci. 28(12), 2959–2964 (2008)
12. Kolmogorov, V., Boykov, Y.: What metrics can be approximated by geo-cuts, or global optimization of length/area and flux. In: IEEE ICCV 2005, vol. 1, pp. 564–571 (2005)
13. Peng, H., Ruan, Z., Long, F., Simpson, J., Myers, E.: V3D enables real-time 3D visualization and quantitative analysis of large-scale biological image data sets. Nature Biotechnology 28(4), 348–353 (2010)
14. Ren, X., Malik, J.: Learning a classification model for segmentation. In: IEEE ICCV 2003, vol. 2, pp. 10–17 (2003)
15. Turaga, S.C., Murray, J.F., Jain, V., Roth, F., Helmstaedter, M., Briggman, K., Denk, W., Seung, H.S.: Convolutional networks can learn to generate affinity graphs for image segmentation. Neural Comput. 22, 511–538 (2010)
16. Turaga, S.C., Briggman, K., Helmstaedter, M., Denk, W., Seung, H.: Maximin affinity learning of image segmentation. In: Bengio, Y., Schuurmans, D., Lafferty, J., Williams, C.K.I., Culotta, A. (eds.) NIPS 2009, pp. 1865–1873 (2009)
17. Vazquez-Reina, A., Miller, E., Pfister, H.: Multiphase geometric couplings for the segmentation of neural processes. In: IEEE CVPR 2009, pp. 2020–2027 (2009)
18. Vu, N., Manjunath, B.: Graph cut segmentation of neuronal structures from transmission electron micrographs. In: IEEE ICIP 2008, pp. 725–728 (2008)
19. Yang, H.-F., Choe, Y.: Electron microscopy image segmentation with graph cuts utilizing estimated symmetric three-dimensional shape prior. In: Bebis, G., Boyle, R., Parvin, B., Koracin, D., Chung, R. (eds.) ISVC 2010. LNCS, vol. 6454, pp. 322–331. Springer, Heidelberg (2010)

Adaptive Energy Selective Active Contour with Shape Priors for Nuclear Segmentation and Gleason Grading of Prostate Cancer

Sahirzeeshan Ali[1], Robert Veltri[2], Jonathan I. Epstein[2], Christhunesa Christudass[2], and Anant Madabhushi[1,⋆]

[1] Rutgers, The State University of New Jersey, New Brunswick, NJ, USA
[2] The Johns Hopkins Hospital, Baltimore, Maryland, USA

Abstract. Shape based active contours have emerged as a natural solution to overlap resolution. However, most of these shape-based methods are computationally expensive. There are instances in an image where no overlapping objects are present and applying these schemes results in significant computational overhead without any accompanying, additional benefit. In this paper we present a novel adaptive active contour scheme (AdACM) that combines boundary and region based energy terms with a shape prior in a multi level set formulation. To reduce the computational overhead, the shape prior term in the variational formulation is only invoked for those instances in the image where overlaps between objects are identified; these overlaps being identified via a contour concavity detection scheme. By not having to invoke all 3 terms (shape, boundary, region) for segmenting every object in the scene, the computational expense of the integrated active contour model is dramatically reduced, a particularly relevant consideration when multiple objects have to be segmented on very large histopathological images. The AdACM was employed for the task of segmenting nuclei on 80 prostate cancer tissue microarray images. Morphological features extracted from these segmentations were found to able to discriminate different Gleason grade patterns with a classification accuracy of 84% via a Support Vector Machine classifier. On average the AdACM model provided 100% savings in computational times compared to a non-optimized hybrid AC model involving a shape prior.

1 Introduction

Active Contours (AC) can be categorized as boundary-based (first generation) and region-based (second generation) schemes [1–3]. Most AC models are not intrinsically capable of handling object occlusion or scene clutter. Therefore, the integration of shape priors into the variational formulation represents a natural way to overcome occlusion. Third generation (hybrid) AC models involve

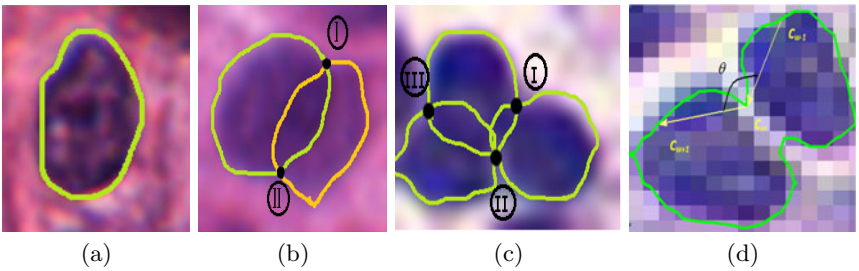
⋆ Thanks to funding agencies: National Cancer Institute (R01CA136535-01, R01CA140772 01, R03CA143991-01), and The Cancer Institute of New Jersey.

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 661–669, 2011.

combining a shape prior with geometric/geodesic active contours that simultaneously achieves registration and segmentation [4–6]. Rousson et al.[5] proposed a novel approach for introducing shape priors into level set representations, focused on 2D closed shapes. A limitation of most third generation AC models, however, is that only one pair of overlapping objects can be accurately resolved at a time. Further, most of these methods are sensitive to model initialization and typically require varying degrees of user intervention [1–6]. Moreover, the efficiency of these hybrid schemes are limited by the computational overhead of the non linearity of the convergence of the evolving curve. Additionally, the shape prior (the most computationally heavy term in the variational formulation) is typically invoked in segmenting every object in the scene, regardless of whether or not an overlap exists. Non-overlapping objects, in most cases, can be segmented by first and second generation AC models alone.

In this paper, a variational adaptive segmentation scheme (AdACM) is presented. AdACM is fully automated and provides concurrent segmentation of all the overlapping and non overlapping objects in the image. Most of the shape based models reported in literature are only able to handle the overlap resolution of a single pair of objects per image. To decrease the user intervention, we propose to automatically initialize our segmentation scheme via the popular Watershed method. To reduce the computational expense of the model, we selectively invoke the shape prior term, only when overlapping objects need to be resolved and segmented. To do this we leverage a technique based on concavity detection [8], to detect the number of overlapping objects, and hence selectively invoke the appropriate energy functionals (see Figure 1). Note that the complexity of the concavity detection scheme [8] is significantly smaller compared to invoking the shape prior.

The Gleason score (obtained by summing the primary and secondary grades of prostate cancer (CaP) in the tissue specimen) is the single most important



(a) (b) (c) (d)

Fig. 1. Selective incorporation of variational terms based on detection of concavity points. In (a) the absence of concavity point reflects the presence of a single nucleus. Similarly (b) and (c) detection of the number of concavity points represents the number of nuclei. (d) Concavity detection: Three consecutive points on s (c_{w-1}, c_w and c_{w+1}) are used to define two vectors (shown with arrows). The angle θ between them is a measure of concavity/convexity of the point c_w [8].

prognostic factor in Cap, determined using the glandular and nuclear architecture and morphology within the tumor [7]. In recent years, computerized image analysis methods have been studied in an effort to overcome the subjectivity of conventional grading system. An important prerequisite to such a computerized CaP grading scheme, however, is the ability to accurately and efficiently segment histological structures (glands and nuclei) of interest. In this work, we leverage the AdACM scheme for automatic segmentation of all nuclei on large digitized tissue microarrays (TMAs) of CaP. Additionally, we leverage previous research that has demonstrated a link between nuclear morphology and gleason grade [7] to develop a nuclear morphology based classifier to predict Gleason grade. The accuracy of this classifier is also implicitly reflective of the performance of AdACM, since accurate nuclear segmentation is a pre-requisite for accurately quantifying nuclear morphology.

2 Hybrid Active Contour Model

The contours that segment the nuclear-boundaries are represented using the level set method, and are evolved by minimizing the variational energy functional. Under the level set framework, the contour is represented implicitly as the zero level of a higher dimensional embedding function, and the contour propagation is performed by evolving the embedding function. This enables handling topological changes of the boundary (splitting and merging) easily.

2.1 Shape Term - F_{shape}

We combine the shape prior, ψ, with a Geodesic Active Contour (GAC) to create the shape functional. ψ, is created using the statistical methods described in [5]. Each shape in the training sample is embedded as the zero level set of a higher dimensional surface. The Signed Distance Function (SDF) is used to encode the distance between the level set (shape contour) and the grid pixels. The level set formulation of the shape functional is expressed as:

$$F_{shape} = \int_{\Omega} (\phi(\mathbf{x}) - \psi(\mathbf{x}))^2 |\nabla\phi| \delta(\phi) d\mathbf{x} \tag{1}$$

where $\{\phi\}$ is a level set function, ψ is the shape prior, $\delta(.)$ is the Dirac function, and $\delta(\phi)$ is the contour measure on $\{\phi = 0\}$. Equation 1 introduces a shape prior in such a way that only objects of interest similar to the shape prior can be recovered, and all unfamiliar image structures are suppressed. It evaluates the shape difference between the level set ϕ and the the shape prior ψ at each iteration of the evolution. However, this formulation only solves for a single level set consistent with the shape prior.

2.2 Region Homogeneity Term

We define a functional to drive the shape functional towards a homogeneous intensity region corresponding to the shape of interest. The functional F_{region}

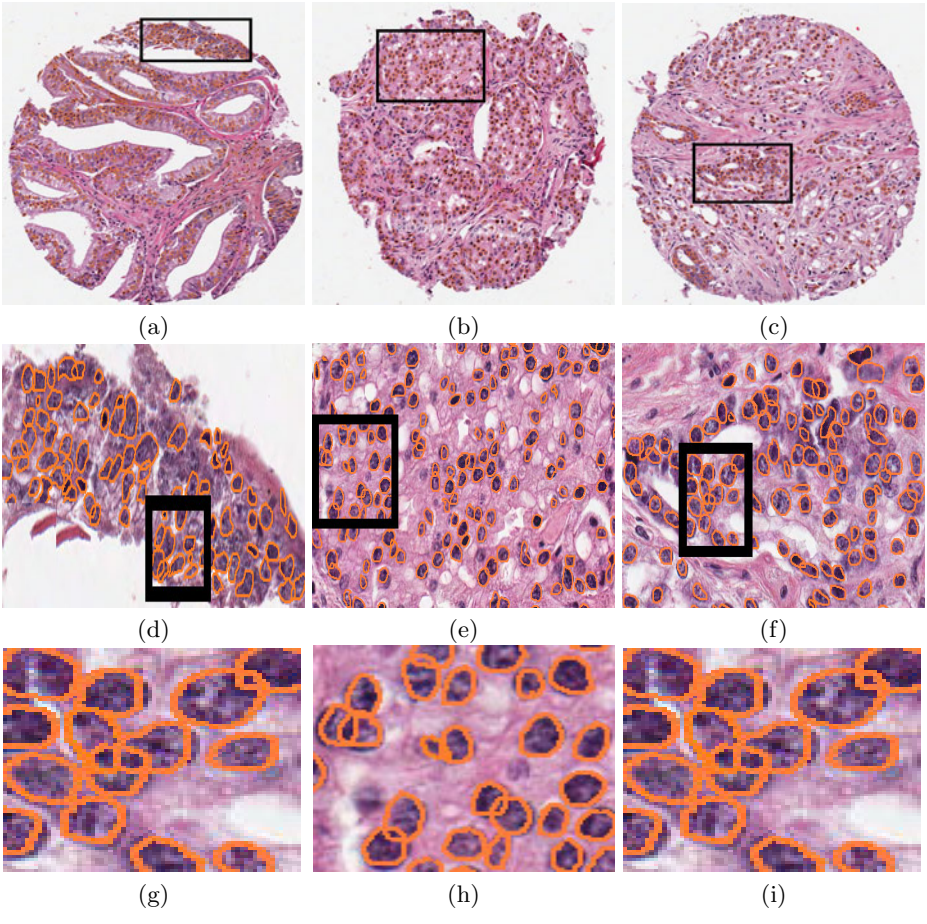


Fig. 2. (a)-(c) Nuclear segmentations (orange boundaries) for three different TMA cylinders corresponding to Gleason scores 6, 7 and 9. (b)-(f) Magnified ROI's from (a)-(c), respectively, reveal that our hybrid ACM (AdACM) with a selective shape prior is able to accurately segment almost all nuclei. (g)-(i) Further magnification of the ROI shown in (a), (b), (c) reveals that our model is able to accurately resolve overlaps and intersections.

can be written with the shape function ψ and statistics of partitioned foreground and background regions, u_{in}, u_{out}:

$$F_{region}(\psi, u_{in}, u_{out}) = \int_{\Omega} \Theta_{in} H_{\psi} d\mathbf{x} + \int_{\Omega} \Theta_{out} H_{-\psi} d\mathbf{x}, \tag{2}$$

where ψ is the shape function, $H(.)$ is the Heaviside function, $\Theta_r = |I - u_r|^2 + \mu|\nabla u_r|^2$ and $r \in \{in, out\}$.

2.3 Combining Shape, Boundary and Region-Based Functionals

We combine F_{shape} and F_{region} into a variational formulation:

$$F = \underbrace{\beta_s \int_{\Omega} (\phi(\mathbf{x}) - \psi(\mathbf{x}))^2 |\nabla\phi| \delta(\phi) d\mathbf{x} +}_{Shape+boundary force} \quad \underbrace{\beta_r \int_{\Omega} \Theta_{in} H_{\psi} d\mathbf{x} + \int_{\Omega} \Theta_{out} H_{-\psi} d\mathbf{x}}_{Region force} \tag{3}$$

where $\beta_s, \beta_r > 0$ are constants that balance contributions of the boundary based shape prior and the region term. This is an extension of the work of Chen et al in [6]. This formulation effectively integrates shape prior with local and regional intensity information into an unified variational formulation.

2.4 Segmenting Multiple Objects under Mutual Occlusion

The level set formulation in Equation (3) is limited in that it allows for segmentation of only a single object at a time. In this work, we incorporate the method presented in [9] into Equation 3. Consider a given image consisting of multiple objects $\{O_1, O_2, \cdots, O_n\}$ of the same shape. For the problems considered in this work (nuclei segmentation on histopathology images), all nuclei are assumed to be roughly elliptical in shape. Instead of partitioning the image domain into mutually exclusive regions, we allow each pixel to be associated with multiple objects or the background. Specifically, we try to find a set of characteristic functions χ_i such that: $\chi_i(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{x} \in O_i \\ 0 & \text{otherwise.} \end{cases}$ We associate one level set per object in such a way that any $O_a, O_b, a, b \in \{1, 2, \cdots, n\}$ are allowed to overlap with each other within the image. These level set components may both be positive on the area of overlap, and enforce the prior on the shapes of objects extracted from the image.

We consider a case of segmenting two objects within an input image. Given an image with two similarly shaped objects $O_a, O_b, a, b \in \{1, \cdots, n\}$, and for simplicity, assume that they are consistent with the shape prior ψ. Then simultaneous segmentation of O_a, O_b with respect to ψ is solved by minimizing the following modified version of Equation (3):

$$\begin{aligned} F(\Phi, \Psi, u_{in}, u_{out}) = & \sum_{i=1}^{N=2} \int_{\Omega} (\phi_i(\mathbf{x}) - \psi(\mathbf{x}))^2 |\nabla\phi_i| \delta(\phi_i) d\mathbf{x} + \beta_r \int_{\Omega} \Theta_{in} H_{\chi_1 \vee \chi_2} d\mathbf{x} \\ & + \int_{\Omega} \Theta_{out} - H_{\chi_1 \vee \chi_2} d\mathbf{x} + \omega \int_{\Omega} H_{\chi_1 \wedge \chi_2} d\mathbf{x} + \sum_{i=1}^{N=2} \int_{\Omega} (\phi_i - \psi_i)^2 d\mathbf{x} \end{aligned} \tag{4}$$

with $H_{\chi_1 \vee \chi_2} = H_{\psi_1} + H_{\psi_2} - H_{\psi_1} H_{\psi_2}, H_{\chi_1 \wedge \chi_2} = H_{\psi_1} H_{\psi_2}$ where $\Phi = (\phi_1, \phi_2)$ and $\Psi = (\psi_1, \psi_2)$. The fourth term penalizes the overlapping area between the two segmenting regions, and prevents the two evolving level set functions from becoming identical. Minimizing Equation 4 iteratively with respect to dynamic variables, yields the associated Euler-Lagrange equations. The above model can be adapted for N objects (proof not shown).

3 Selectively Invoking Energy Terms in Hybrid ACM

3.1 Watershed Based Initialization and Concavity Detection

We use the popular watershed transformation to obtain the initial delineations of nuclear boundaries in the entire image. By creating the binary mask of the delineations, we obtain the estimated boundaries of the nuclei present.

High concavity points are characteristic of contours that enclose multiple objects and represent junctions where object intersection occurs (Figure 1(d)). The area $\mathcal{A}(s)$ of the closed sub-contour s is compared to predetermined area of an ideal nucleus τ_A (empirically set to $\tau_A = 33$). Hence a sub-contour is eligible for a split if $\mathcal{A}(s) > \tau_A$. Since $c = (x, y)$, the difference between any two points c_w and c_{w-1} will represent a vector in 2D. Concavity points are detected by computing the angle between vectors defined by three consecutive points $(c_{w-1}, c_w, c_{w+1}) \in s$. The degree of concavity/convexity is proportional to the angle $\theta(c_w)$ as shown in Figure 1(d). $\theta(c_w)$ can be computed from the dot product relation (Equation 5):

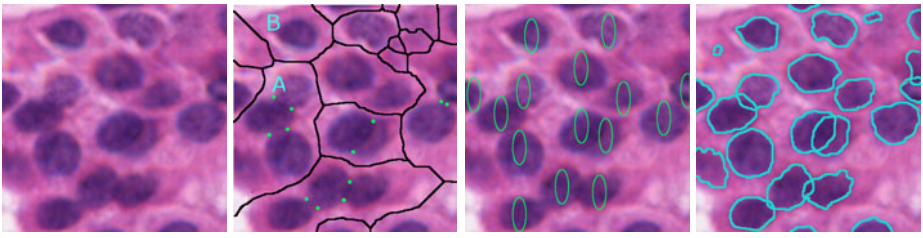
$$\theta(c_w) = \pi - \arccos\left(\frac{(c_w - c_{w-1}) \cdot (c_{w+1} - c_w)}{||(c_w - c_{w-1})|| \;\; ||(c_{w+1} - c_w)||}\right). \tag{5}$$

3.2 Adaptive Selection of Energy Functionals in Hybrid AC Model

Number of detected concavity points, $c_w \leq 1$, indicates presence of a single non overlapping nucleus. In such cases, shape prior constraint is not necessary and we reduce the model to only employ the region term by setting $\beta_s = 0$. Similarly, l number of c_w indicate the presence of l overlapping objects. Hence in those regions we initialize the model with the integrated hybrid model (region, boundary, shape terms) with l level sets and set $N = l$ (in Equation 4). Figure 3 illustrates the work flow from initialization to final segmentation for AdACM.

4 Experimental Design and Results

Since it is not feasible to evaluate AdACM quantitatively on a per nucleus basis (manual annotation not practical for thousands of nuclei), AdACM was instead evaluated in terms of the ability of the morphologic descriptors extracted from the nuclear boundaries (based off [7]) to distinguish different Gleason patterns of CaP from a total of 40 (2 images per study) TMA images obtained from prostate needle core biopsy samples. The 40 studies comprised 13 Gleason patterns 6 (3+3), 8 pattern 7 (3+4), 7 pattern 8 (4+4) and 5 pattern 9 (4+5) studies where the first number in the parenthesis refers to the primary and the second number to the secondary Gleason grade. Additionally the ability of the model to selectively invoke energy terms in the variational functional was also evaluated in terms of the (a) overlaps between nuclei resolved (for a randomly selected subset of nuclei) and (b) computational savings over a hybrid model without selective invocation of the shape term.



(a) (b) (c) (d)

Fig. 3. Constituent modules of the AdACM. (a) Original image; (b) Watershed segmentation of individual nuclei with an overlay detected concavity points; (c) Placement of initial level sets in the image; (d) final segmentation. Note that in the region shown as A where 4 concavity points where detected, 3 level sets were initialized with the shape prior, whereas in region B, only a single level set (only region based term) was initialized.

4.1 Discriminating Gleason Patterns Based off Nuclear Morphology

A total of 7 nuclear features from each of the segmented nuclei were extracted (Morphologic features include: Area Overlap Ratio, Average Radial Ratio, Compactness, Convexity, Mean Nuclear Area, Mean Nuclear Perimeter, Mean Nuclear Diameter). We apply PCA to this feature set to visualize the arrangement of different Gleason patterns in the reduced embedding space. Figure 4(a) illustrates the PCA representation of the 7 morphologic features averaged over each of 40 studies (total of 80 images) and reveals a clear separation between Gleason patterns 3 and 4. Similarly, by exposing the labels for the Gleason scores for each of 40 studies (Figure 4(b)), one can appreciate the separation between Gleason patterns 6, 7, 8, and 9. Note that the PCA plots suggest that the nuclear shape features are able to capture the subtle morphologic differences between the different Gleason patterns, in turn reflecting the accuracy of AdACM. The separation of the intermediate primary grade 3 and grade 4 tumors in the reduced PCA space was also quantitatively evaluated using a support vector machine (SVM) classifier. For a database of 80 images, the SVM classifier achieved an accuracy of $83.8 \pm 0.4\%$ in distinguishing primary grade 3 from 4. Classifier evaluation was done via a 3 fold cross validation scheme, over 10 successive runs.

4.2 Evaluating Overlap Resolution and Segmentation Accuracy

Quantitative comparison of AdACM with a GAC (Geodesic Active Contour) [2] and the Rousson-Deriche shape based model (RD) [5] was performed (see Table 1). Overlap resolution was evaluated via: $OR = \frac{\#\text{ overlaps resolved}}{\text{Total }\#\text{ of overlaps}}$.

4.3 Computational Efficiency

We evaluated the computational efficiency of AdACM with respect to a hybrid ACM (HACM) which did not employ selective invocation of the shape prior.

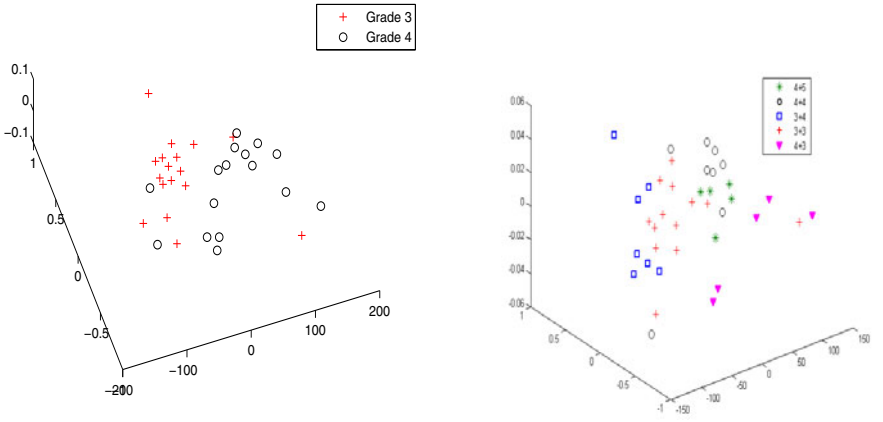


Fig. 4. PCA representation of nuclear morphologic features reveals separation of (a) primary grade 3 and grade 4 and (b) Gleason patterns 6, 7, 8 and 9

Table 1. Quantitative evaluation of segmentation, and Overlap Resolution between the Geodesic Active Contour [2], Rousson-Deriche [5] and AdACM for 200 randomly selected nuclei across the 80 TMA images

	Sensitivity	Specificity	Mean Average Distance	OR
GAC	0.28	0.92	7.1	0.04
RD	0.76	0.88	1.5	0.74
AdACM	**0.82**	**1.0**	**1.1**	**0.90**

On image patches of 200 × 200 pixels and in the presence of 120 nuclei with 40 overlaps, AdACM required 250s to accurately segment all objects and resolve all intersections, compared to HACM which took 550s; all evaluations being performed on a 3 GHz, dual core processor with 4 GB RAM.

5 Concluding Remarks

In this work we presented a new hybrid active contour that employs a shape prior for concurrent segmentation of multiple nuclei. Our model (AdACM) selectively invokes the shape prior terms (computationally expensive step) in those image regions where objects overlap, overlaps being determined via a concavity detection scheme. This selective invocation of energy terms in the variational formulation yields a hybrid ACM that is both accurate and computationally efficient. Morphologic features derived from nuclei segmented via AdACM enabled discrimination of different Gleason grade patterns, on prostate cancer TMAs, reflecting the segmentation accuracy of AdACM. Additionally, independent segmentation

based performance based measures also reflected the efficacy of our scheme. In future work we intend to leverage AdACM in the context of nuclear and cell segmentation in other domains within digital pathology.

References

1. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: active contour models. International J. of Computer Vision, 321–331 (1987)
2. Caselles, V., Kimmel, R., Sapiro, G.: Geodesic active contours. Int.J. Comput. Vision 22(1), 61–79 (1997)
3. Chan, T., Vese, L.: Active contours without edges. IEEE Trans. on Image Processing 10(2), 266–277 (2001)
4. Fang, W., Chan, K.: Statistical Shape Influence in Geodesic Active Contours. IEEE CVPR 40(8), 2163–2172 (2007)
5. Rousson, M., Paragios, N.: Shape priors for level set representations. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002. LNCS, vol. 2351, pp. 78–92. Springer, Heidelberg (2002)
6. Chan, T.: Level set based shape prior segmentation. IEEE CVPR 2, 1164–1170 (2005)
7. Veltri, R., Isharwal, S., Mille, M., Epstein, J.I., Partin, A.: Nuclear Roundness Variance Predicts Prostate Cancer Progression,Metastasis, and Death: A Prospective EvaluationWith up to 25 Years of Follow-Up After Radical Prostatectomy. The Prostate 70, 1333–1339 (2010)
8. Fatakdawala, H., Xu, J., Basavanhally, A., Bhanot, G., Ganesan, S., Feldman, M., Tomaszewski, J., Madabhushi, A.: Expectation Maximization driven Geodesic Active Contour with Overlap Resolution (EMaGACOR): Application to Lymphocyte Segmentation on Breast Cancer Histopathology. IEEE TBME 57(7), 1676–1689 (2010)
9. Zhang, Q., Pless, R.: Segmenting multiple familiar objects under mutual occlusion. In: ICIP (2006)

Detection of Neuron Membranes in Electron Microscopy Images Using Multi-scale Context and Radon-Like Features

Mojtaba Seyedhosseini[1,2,⋆], Ritwik Kumar[3], Elizabeth Jurrus[2], Rick Giuly[4], Mark Ellisman[4], Hanspeter Pfister[5], and Tolga Tasdizen[1,2]

[1] Electrical and Computer Engineering Department, University of Utah
[2] Scientific Computing and Imaging Institute, University of Utah
mseyed@sci.utah.edu
[3] IBM Almaden Research Center, San Jose
[4] National Center for Microscopy and Imaging Research, University of California, San Diego
[5] School of Engineering and Applied Sciences, Harvard University

Abstract. Automated neural circuit reconstruction through electron microscopy (EM) images is a challenging problem. In this paper, we present a novel method that exploits multi-scale contextual information together with *Radon-like features* (RLF) to learn a series of discriminative models. The main idea is to build a framework which is capable of extracting information about cell membranes from a large contextual area of an EM image in a computationally efficient way. Toward this goal, we extract RLF that can be computed efficiently from the input image and generate a scale-space representation of the *context images* that are obtained at the output of each discriminative model in the series. Compared to a single-scale model, the use of a multi-scale representation of the context image gives the subsequent classifiers access to a larger contextual area in an effective way. Our strategy is general and independent of the classifier and has the potential to be used in any context based framework. We demonstrate that our method outperforms the state-of-the-art algorithms in detection of neuron membranes in EM images.

Keywords: Machine learning, Membrane detection, Neural circuit reconstruction, Multi-scale context, Radon-like features (RLF).

1 Introduction

Electron microscopy (EM) is an imaging technique that can generate nanoscale images that contain enough details for reconstruction of the connectome, i.e., the wiring diagram of neural processes in the mammalian nervous system [4,11]. Because of the large number and size of images, their manual analysis is infeasible and in some cases may take more than a decade [3]. Hence, automated image analysis is required. However, fully automatic reconstruction of the connectome

⋆ Corresponding author.

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 670–677, 2011.

is challenging because of the complex intracellular structures, noisy texture, and the large variation in the physical topologies of cells [5]. Therefore, a successful automated method must overcome these issues in order to reconstruct the neural circuit with high accuracy.

Many supervised and unsupervised techniques have been proposed to solve the connectome reconstruction problem. Macke *et al.* [9] proposed a contour propagation model that minimizes an energy function to find the cell membranes. However, this active contour model can get stuck in local minima due to the complex intracellular structures and may find false boundaries [10]. Vu and Manjunath [12] proposed a graph-cut framework that minimizes an energy defined over the image intensity and the intensity gradient field. But, the graph-cut method might be misled by the complex intracellular structure of the EM images and requires the user to correct segmentation errors. Kumar *et al.* [7] introduced a set of so-called Radon-like features (RLF), which take into account both texture and geometric information and overcome the problem of complex intracellular structures but only achieve modest accuracy levels due to the lack of a supervised classification scheme.

Supervised methods that use contextual information [2] have been proven successful to solve the reconstruction problem. Jain *et al.* [5] proposed a convolutional neural network for restoring membranes in EM images. Convolutional networks take advantage of context information from increasingly large regions as one progresses through the layers. To capture context from a large region, however, convolutional networks need many hidden layers, adding significant complexity to training. Jurrus *et al.* [6] proposed a framework to detect neuron membranes that integrates information from the original image together with contextual information by learning a series of artificial neural networks (ANN). This makes the network much easier to train because the classifiers in the series are trained one at a time and in sequential order.

Even though these approaches improve the accuracy of the segmentation over unsupervised methods, they don't utilize the context information in an effective way. In [6], Jurrus *et al.* utilize context locations that are selected by a stencil and use them as input to a neural network. The performance of the classifier can be improved by using context from a large neighborhood; however, it is not practical to sample every pixel in a very large context area because of computational complexity and the overfitting problem. To address this problem, we develop a multi-scale strategy to take advantage of context from a larger area while keeping the computational complexity tractable and avoiding overfitting. We apply a series of linear averaging filters to the context image consecutively to generate a scale-space representation [1] of the context. Thus the classifier can have as input a small neighborhood, i.e., a 5×5 patch, at the original scale as well as the coarser scales. While scale-space methods are well known, to our knowledge their use for modelling context in classification problems is novel. Combining scale-space representation and contextual information leads to a novel segmentation framework that provides more information from the context for the classifiers in the series. This extra information from the context

helps the later classifiers to correct the mistakes of the early stages and thus improves the overall performance.

In addition to the above problem with existing context based methods that we address in this paper, we also note that none of the existing methods make use of textural and geometric features specifically designed for connectome images. We also address this by incorporating the recently proposed Radon-like features [7] in our method. RLF, which can be efficiently computed, provide our classifier discriminative information in addition to that present in the grayscale micrograph. It must be emphasized that [7] proposes that RLF be used only at a single scale with certain set of parameters. We sidestep this parameter tuning problem by computing RLF at various scales and using them all in our classifier.

2 Sequential Training with Context

Given a set of training images and corresponding groundtruth labels for each pixel, we learn a set of classifiers in sequential order as in [6]. The first classifier is trained only on the input image features. The output of this classifier, the probability image map, is used together with the input image features to train the next stage classifier. The algorithm iterates until the improvement in the performance of the current stage is small compared to the previous stage.

Let $X = (x(i,j))$ be the input image that comes with a ground truth $Y = (y(i,j))$ where $y(i,j) \in \{-1, 1\}$ is the class label for pixel (i,j). The training set is $T = \{(X_k, Y_k); k = 1, \ldots, M\}$ where M denotes the number of training images. A typical approximation of the MAP estimator for Y given X is obtained by using the Markov assumption that decreases the computational complexity:

$$\hat{y}_{MAP}(i,j) = \operatorname{argmax} p(y(i,j)|X_{N(i,j)}), \tag{1}$$

where $N(i,j)$ denotes all the pixels in the neighborhood of pixel (i,j). Instead of using the entire input image the classifier has access to a limited number of neighborhood pixels at each input pixel (i,j).

In the series-ANN [6], a classifier is trained based on the neighborhood features at each pixel. We call the output image of this classifier $C = (c(i,j))$. The next classifier is trained not only on the neighborhood features of X but also on the neighborhood features of C. The MAP estimation for this classifier is:

$$\hat{y}_{MAP}(i,j) = \operatorname{argmax} p(y(i,j)|X_{N(i,j)}, C_{N'(i,j)}), \tag{2}$$

where $N'(i,j)$ is the set of all neighborhood pixels of pixel (i,j) in the context image. Note that N and N' can be different neighborhoods. The same procedure is repeated through the different stages of the series classifier until convergence. It is worth mentioning that Eq. 2 is closely related to the CRF model [8]; however in our approach multiple models in series are learned, which is an important difference from standard CRF approaches.

According to Eq. 2, context provides prior information to solve the MAP problem. Even though the Markov assumption is reasonable and makes the problem

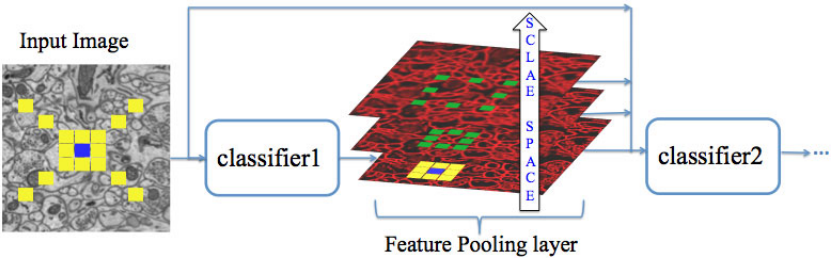


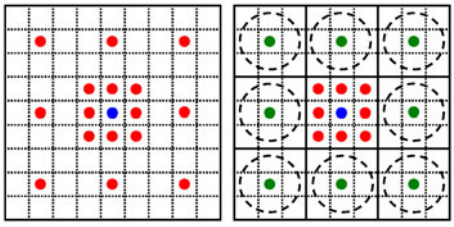
Fig. 1. Illustration of the multi-scale contextual model. Each feature map is sampled at different scales (green rectangles). The blue rectangles represent the center pixel and the yellow rectangles show the selected context locations at original scale.

tractable, it still results in a significant loss of information from global context. However, it is not practical to sample every pixel in a very large neighborhood area of the context due to the computational complexity problem and overfitting. Previous approaches [6] have used a sparse sampling approach to cover large context areas as shown in Fig. 2(a). However, single pixel contextual information at the finest scale conveys only partial information about its neighborhood in a sparse sampling strategy while each pixel at the coarser scales conveys more information about its surrounding area due to the use of averaging filters. Furthermore, single pixel context is noise prone whereas context at coarser scales is more robust due to the averaging. In other words, while it is reasonable to sample context at the finest level at a distance of a few pixels, sampling context at the finest scale tens to hundreds of pixels away is error prone and presents a non-optimal summary of its local area. We argue that more information can be obtained by creating a scale-space representation of the context and allowing the classifier access to samples of small patches at each scale. Conceptually, sampling from scale-space representation increases the effective size of the neighborhood while keeping the number of samples small.

3 Multi-scale Contextual Model

Multi-scale contextual model is shown in Fig. 1. Each stage is composed of two layers: a classifier layer and a feature pooling layer. **Classifier:** Different types of classifiers can be used in series architecture such as AdaBoost and neural networks. The first classifier operates only on the input image while the later stages are trained on both the input image and the context from the previous stage. **Feature Pooling:** In the conventional series structure, the feature pooling layer simply takes sparsely sampled context as in Fig. 2(a) and combines them with input image features. In the proposed method, the feature pooling layer treats each feature map as an image and creates a scale-space representation by applying a series of Gaussian filters. This results in a feature map with lower resolution that is robust against the small variations in the location of features and noise.

Fig. 2 shows our sampling strategy versus single space sampling strategy. In Fig. 2(b) the classifier can have as an input the center 3×3 patch at the original scale and a summary of 8 surrounding 3×3 patches at a coarser scale. The green circles in Fig. 2(b) are more informative and less noisy compared to their equivalent red circles in Fig. 2(a). The summaries become more informative as the number of scales increases. For example, in the first scale the summary is computed over 9 pixels (3×3 neighborhood) while it is computed over 25 pixels (5×5 neighborhood) in the second scale. In practice, we use Gaussian averaging filters to create the summary (green circles in Fig. 2(b)). Other methods like max-pooling can be used instead of Gaussian averaging. The number of scales and the Gaussian filter size are set according to the application characteristics.



(a) (b)

Fig. 2. Sampling strategy of context: Sampling at (a) single-scale (b) multi-scale. Green circles illustrate the summary of pixels in dashed circles.

Taking multiple scales into account, Eq. 2 can be rewritten as:

$$\hat{y}_{MAP}(i,j) = \text{argmax}\, p(y(i,j)|X_{N(i,j)}, C_{N_0'(i,j)}(0), \ldots, C_{N_l'(i,j)}(l)), \qquad (3)$$

where $C(0), \ldots, C(l)$ denote the scale-space representation of the context and $N_0^{'}(i,j), \ldots, N_l^{'}(i,j)$ are corresponding sampling structures. Unlike Eq. 2 that uses the context in a single scale, Eq. 3 takes advantage of multi-scale contextual information. Although in Eq. 3 we still use the Markov assumption, the size of the neighborhood is larger, and thus we lose less information compared to Eq. 2.

4 Radon-Like Features

As mentioned earlier, the overall performance of our method can be improved by extracting RLF from the input image in addition to pixel intensities. It has been shown empirically that trying to segment the structures in connectome images using only geometric or textural features is not very effective [7]. RLF were proposed as a remedy to this problem as they are designed to leverage both the texture and the geometric information present in the connectome images to segment structures of interest. As a first step, RLF use the edge map of a connectome image as a means to divide it into regions that are defined by the geometry of the constituent structures. Next, for each pixel, line segments with their end points on the closest edges are computed in all directions. Finally, for each pixel, a scalar value is computed along each direction using the information in the original image along these line segments using a so-called extraction function. Extraction functions tuned to extract cell boundaries, mitochondria, vesicles, and cell background have been defined in [7].

In this paper, we are interested in obtaining the cell boundaries from the connectome images. Moreover, we intend to define a supervised scheme to automatically segment the cell boundaries while [7] presented an unsupervised,

and consequently less accurate, framework. Both of these objectives allow us to use the RLF in a more targeted manner towards cell boundary segmentation. Foremost, we use not just the cell boundary extraction function but also the mitochondria extraction function since we train our classifier to not select mitochondria boundaries as cell boundaries. Secondly, we use what we call *multi-scale RLF* by computing RLF at multiple scales and for different edge threshold settings. This richer set of features allow for correct detection of cell boundaries in the regions that cannot be detected by the original RLF as proposed in [7] and avoids the need for extensive parameter tuning.

Combining these set of features and the multi-scale contextual model, the update equation for the framework can be written as:

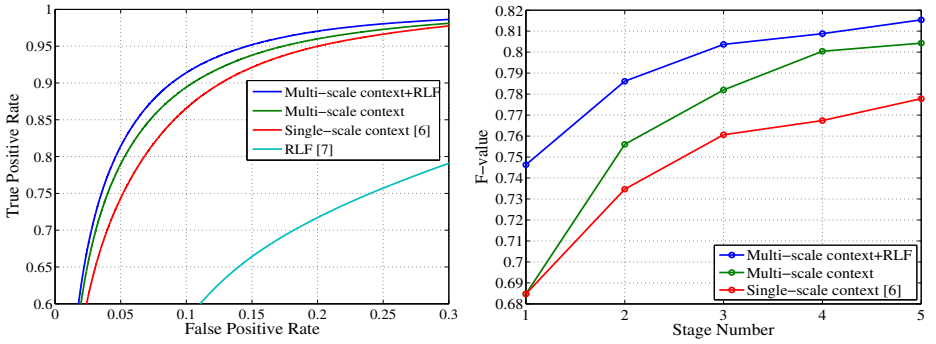
$$\hat{y}_{MAP}^{k+1}(i,j) = \operatorname{argmax} p(y(i,j)|X_{N(i,j)}, f(X_{N(i,j)}), \\ C^k_{N'_0(i,j)}(0), \ldots, C^k_{N'_l(i,j)}(l)), \tag{4}$$

where $C^k(0), \ldots, C^k(l)$ are the scale-space representation of the output of classifier stage k, $k = 1, \ldots, K-1$, $\hat{y}_{MAP}^{k+1}(i,j)$ is the output of the stage $k+1$ and $f(.)$ is the RLF function. In turn, the $k+1$'st classifier output as defined in Eq. 4 creates the context for the $k+2$'nd classifier. The model repeats Eq. 4 until the performance improvement between two consecutive stages becomes small.

5 Experimental Results

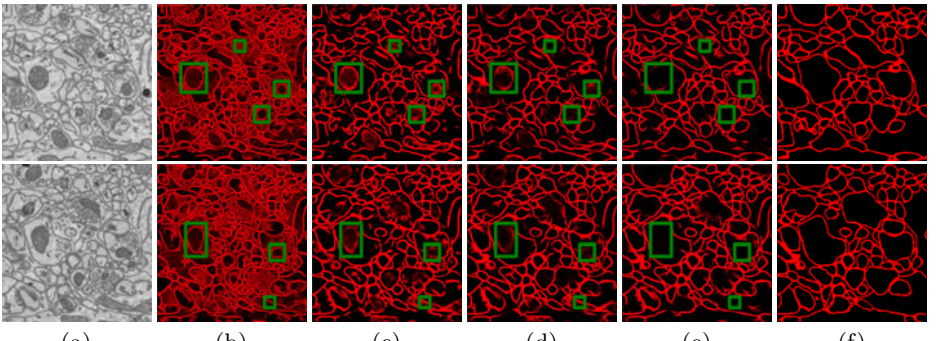
We test the performance of our proposed method on a set of 70 EM images of a mouse cerebellum with corresponding groundtruth maps. The groundtruth images were annotated by an expert who marked neuron membranes with a one-pixel wide contour. 14 of these images were used for training and the remaining images were used for testing. In this experiment, we employed MLP-ANNs as the classifier in a series structure, as in [6]. Each MLP-ANN in the series had one hidden layer with 10 nodes.

To optimize the network performance, 5.5 million pixels were randomly selected from the training images such that there are twice the number of negative examples, than positive as in [6]. Input image feature vectors were computed on a 11×11 stencil centered on each pixel. The same stencil was used to sample the RLF for cell boundaries (at two scales) and mitochondria. The context features were computed using 5×5 patches at four scales (one at original resolution and three at coarser scales). The classifier then gets as input the 5×5 patch at the original resolution ($C_{N'_0(i,j)}(0)$) and 5×5 patches at three coarser scales ($C_{N'_l(i,j)}(l)$). The ROC curves for pixel-wise membrane detection are shown in Fig. 3(a). It can be noted that our method outperforms the state-of-the-art methods proposed in [6] and [7]. The average $F - value = \frac{2 \times Precision \times Recall}{Precision + Recall}$ at zero threshold for different stages and different methods is shown in Fig. 3(b). The performance of the multi-scale contextual model without RLF is 2.65% better than using a single-scale context [6]. This improvement increases to 3.76% when we use RLF in addition to multi-scale contextual information. Fig. 4 shows some



(a) (b)

Fig. 3. (a) The ROC curves for test images and for different methods. (b) The F-value at different stages for different methods. The F-value for RLF method [7] is 59.40%.



(a) (b) (c) (d) (e) (f)

Fig. 4. Test results for the membrane detection for two different input images: (a) Input image, the remaining columns show the output results (probability maps) for (b) RLF [7] (c) single-scale context [6] (d) multi-scale context (e) multi-scale context+RLF, and (f) shows the manually marked groundtruth

examples of our test images and corresponding membrane detection results for different methods. As shown in our results, the approach presented here performs better in membrane detection compared to [6], and it is more successful in removing undesired parts (green rectangles) from inside cells.

6 Conclusion

This paper introduced an image segmentation algorithm using a multi-scale contextual model. The main idea of our method is to take advantage of context images at different scales instead of a single scale, thereby providing the classifier with a richer set of information. We also modified the RLF to extract more

information from different structures of the input image. The proposed method is very general and does not depend on any particular classifier or any specific scale-space method.

We applied our method to membrane detection in EM images. Results indicate that the proposed method outperforms state-of-the-art algorithms while maintaining nearly identical computational complexity. We used linear averaging filters to generate the scale-space representation of the context. In future work, we will conduct a full study of the effect of scale-space depth and the advantage of using other linear or nonlinear scale-space methods.

Acknowledgements. This work was supported by NIH 1R01NS075314-01 (TT,MHE), NSF PHY-0835713 (HP), and NIH P41 RR004050 (MHE).

References

1. Babaud, J., Witkin, A.P., Baudin, M., Duda, R.O.: Uniqueness of the gaussian kernel for scale-space filtering. IEEE Transactions on PAMI 8(1), 26–33 (1986)
2. Belongie, S., Malik, J., Puzicha, J.: Shape matching and object recognition using shape contexts. IEEE Transactions on PAMI 24(4), 509–522 (2002)
3. Briggman, K.L., Denk, W.: The human connectome: a structural description of the human brain. Current Opinion in Neurobiology 16(5), 562–570 (2006)
4. Denk, W., Horstmann, H.: Serial block-face scanning electron microscopy to reconstruct three-dimensional tissue nanostructure. PLoS Biology 2, e329 (2004)
5. Jain, V., Murray, J.F., Roth, F., Turaga, S., Zhigulin, V., Briggman, K.L., Helmstaedter, M.N., Denk, W., Seung, H.: Supervised learning of image restoration with convolutional networks. In: Proceedings of ICCV, pp. 1–8 (2007)
6. Jurrus, E., Paiva, A.R.C., Watanabe, S., Anderson, J.R., Jones, B.W., Whitaker, R.T., Jorgensen, E.M., Marc, R.E., Tasdizen, T.: Detection of neuron membranes in electron microscopy images using a serial neural network architecture. Medical Image Analysis 14(6), 770–783 (2010)
7. Kumar, R., Va andzquez Reina, A., Pfister, H.: Radon-like features and their application to connectomics. In: IEEE Computer Society Conference on CVPRW, pp. 186–193 (June 2010)
8. Lafferty, J., McCallum, A., Pereira, F.: Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In: Proceedings of ICML, pp. 282–289 (2001)
9. Macke, J.H., Maack, N., Gupta, R., Denk, W., Schlkopf, B., Borst, A.: Contour-propagation algorithms for semi-automated reconstruction of neural processes. Journal of Neuroscience Methods 167(2), 349–357 (2008)
10. Mishchenko, Y.: Automation of 3d reconstruction of neural tissue from large volume of conventional serial section transmission electron micrographs. Journal of Neuroscience Methods 176(2), 276–289 (2009)
11. Sporns, O., Tononi, G., Ktter, R.: The human connectome: a structural description of the human brain. PLoS Computational Biology 1, e42 (2005)
12. Vu, N., Manjunath, B.S.: Graph cut segmentation of neuronal structures from transmission electron micrographs. In: Proceedings of ICIP, pp. 725–728 (2008)

Author Index