

Gabor Fichtinger
Anne Martel
Terry Peters (Eds.)

LNCS 6893

Medical Image Computing and Computer-Assisted Intervention – MICCAI 2011

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

3
Part III



MICCAI

 Springer

Commenced Publication in 1973

Founding and Former Series Editors:

Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Editorial Board

David Hutchison

Lancaster University, UK

Takeo Kanade

Carnegie Mellon University, Pittsburgh, PA, USA

Josef Kittler

University of Surrey, Guildford, UK

Jon M. Kleinberg

Cornell University, Ithaca, NY, USA

Alfred Kobsa

University of California, Irvine, CA, USA

Friedemann Mattern

ETH Zurich, Switzerland

John C. Mitchell

Stanford University, CA, USA

Moni Naor

Weizmann Institute of Science, Rehovot, Israel

Oscar Nierstrasz

University of Bern, Switzerland

C. Pandu Rangan

Indian Institute of Technology, Madras, India

Bernhard Steffen

TU Dortmund University, Germany

Madhu Sudan

Microsoft Research, Cambridge, MA, USA

Demetri Terzopoulos

University of California, Los Angeles, CA, USA

Doug Tygar

University of California, Berkeley, CA, USA

Gerhard Weikum

Max Planck Institute for Informatics, Saarbruecken, Germany

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 III

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
ISBN 978-3-642-23625-9
DOI 10.1007/978-3-642-23626-6
Springer Heidelberg Dordrecht London New York

e-ISSN 1611-3349
e-ISBN 978-3-642-23626-6

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

© Springer-Verlag Berlin Heidelberg 2011

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

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 15th 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
 Elsa Angelini
 Stephen Aylward
 Christian Barillot
 Wolfgang Birkfellner
 Albert C.S. Chung

Ela Claridge
 D. Louis Collins
 Dorin Comaniciu
 Christos Davatzikos
 Marleen de Bruijne
 Hervé Delingette
 James S. Duncan
 Philip (Eddie) Edwards
 Bernd Fischer
 P. Thomas Fletcher
 Alejandro Frangi
 Polina Golland
 Nobuhiko Hata
 David Hawkes
 David Holmes
 Ameet Jain
 Pierre Jannin
 Leo Joskowicz
 Ioannis Kakadiaris
 Rasmus Larsen
 Anant Madabhushi
 Frederik Maes
 Kensaku Mori
 Parvin Mousavi
 Nassir Navab
 Poul Nielsen
 Wiro Niessen
 Sebastien Ourselin
 Nikos Paragios
 Xavier Pennec
 Josien Pluim
 Daniel Rueckert
 Ichiro Sakuma
 Tim Salcudean
 Julia A. Schnabel
 Dinggang Shen
 John G. Sled

University of British Columbia, Canada
 Telecom Paris, France
 Kitware, USA
 INRIA, France
 Medical University of Vienna, Austria
 Hong Kong University of Science and
 Technology, China
 University of Birmingham, UK
 McGill University, Canada
 Siemens Medical, USA
 University of Pennsylvania, USA
 Erasmus Medical College, The Netherlands
 INRIA, France
 Yale University, USA
 Imperial College London, UK
 University of Lübeck, Germany
 University of Utah, USA
 Pompeu Fabra University, Spain
 Massachusetts Institute of Technology, USA
 Harvard Medical School, USA
 University College London, UK
 Mayo Clinic, USA
 Philips Medical Systems, USA
 INRIA, France
 Hebrew University of Jerusalem, Israel
 University of Houston, USA
 Technical University of Denmark
 Rutgers University, USA
 University of Leuven, Belgium
 Nagoya University, Japan
 Queen's University, Canada
 Technical University of Munich, Germany
 University of Auckland, New Zealand
 Erasmus Medical College, The Netherlands
 University College London, UK
 Ecole Central Paris, France
 INRIA, France
 Utrecht University, The Netherlands
 Imperial College London, UK
 University of Tokyo, Japan
 University of British Colombia, Canada
 University of Oxford, UK
 University of North Carolina, USA
 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, Jung-ha	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
Bergholdt, 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, Djamel
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
Krugger, 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
Lingurar, 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, Stanislas
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
Quelleg, Gwenole
Radeva, Petia
Rajagopal, Vijayaraghavan
Ramamurthi, Krishnakumar
Ramezani, Mahdi
Rasouliau, 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
Szczërba, 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
Vanderghenst, 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
 Zwigelaar, 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.

Rogério Richa, for the article entitled: “Robust 3D Visual Tracking for Robotic-Assisted Cardiac Interventions,” co-authored by: Rogério 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”

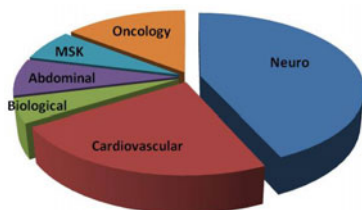
Functional and
Diffusion-Weighted MRI

Anthony J. Sherbondy

“MicroTrack: An Algorithm for Concurrent Projectome and Microstructure Estimation”

Accepted MICCAI 2011 Papers

by Clinical Theme:



by Technical Theme:

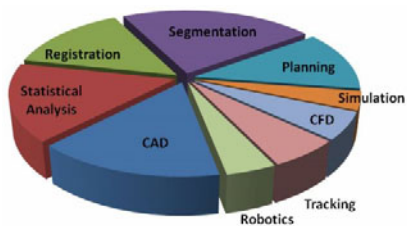


Table of Contents – Part III

Computer Aided Diagnosis and Machine Learning I

Sliding Window and Regression Based Cup Detection in Digital Fundus Images for Glaucoma Diagnosis	1
<i>Yanwu Xu, Dong Xu, Stephen Lin, Jiang Liu, Jun Cheng, Carol Y. Cheung, Tin Aung, and Tien Yin Wong</i>	
Learning from Only Positive and Unlabeled Data to Detect Lesions in Vascular CT Images	9
<i>Maria A. Zuluaga, Don Hush, Edgar J.F. Delgado Leyton, Marcela Hernández Hoyos, and Maciej Orkisz</i>	
Regularized Tensor Factorization for Multi-Modality Medical Image Classification	17
<i>Nematollah Batmanghelich, Aoyan Dong, Ben Taskar, and Christos Davatzikos</i>	
Detection, Grading and Classification of Coronary Stenoses in Computed Tomography Angiography	25
<i>B. Michael Kelm, Sushil Mittal, Yefeng Zheng, Alexey Tsymbal, Dominik Bernhardt, Fernando Vega-Higuera, S. Kevin Zhou, Peter Meer, and Dorin Comaniciu</i>	
Aggregated Distance Metric Learning (ADM) for Image Classification in Presence of Limited Training Data	33
<i>Gaoyu Xiao and Anant Madabhushi</i>	
Sparse Classification for Computer Aided Diagnosis Using Learned Dictionaries	41
<i>Meizhu Liu, Le Lu, Xiaojing Ye, Shipeng Yu, and Marcos Salganicoff</i>	
A Discriminative-Generative Model for Detecting Intravenous Contrast in CT Images	49
<i>Antonio Criminisi, Krishna Juluru, and Sayan Pathak</i>	
Classification of Alzheimer’s Disease Using a Self-Smoothing Operator	58
<i>Juan Eugenio Iglesias, Jiayan Jiang, Cheng-Yi Liu, Zhuowen Tu, and ADNI</i>	
An Analysis of Scale and Rotation Invariance in the Bag-of-Features Method for Histopathological Image Classification	66
<i>S. Hussain Raza, R. Mitchell Parry, Richard A. Moffitt, Andrew N. Young, and May D. Wang</i>	

Robust Large Scale Prone-Supine Polyp Matching Using Local Features: A Metric Learning Approach	75
<i>Meizhu Liu, Le Lu, Jinbo Bi, Vikas Raykar, Matthias Wolf, and Marcos Salganicoff</i>	
Targeted Optical Biopsies for Surveillance Endoscopies	83
<i>Selen Atasoy, Diana Mateus, Alexander Meining, Guang-Zhong Yang, and Nassir Navab</i>	
Focal Biologically Inspired Feature for Glaucoma Type Classification . . .	91
<i>Jun Cheng, Dacheng Tao, Jiang Liu, Damon Wing Kee Wong, Beng Hai Lee, Mani Baskaran, Tien Yin Wong, and Tin Aung</i>	
The Relevance Voxel Machine (RVoxM): A Bayesian Method for Image-Based Prediction	99
<i>Mert R. Sabuncu and Koen Van Leemput</i>	
Assessment of Regional Myocardial Function via Statistical Features in MR Images	107
<i>Mariam Afshin, Ismail Ben Ayed, Kumaradevan Punithakumar, Max W.K. Law, Ali Islam, Aashish Goela, Ian Ross, Terry Peters, and Shuo Li</i>	

Computer Aided Diagnosis and Machine Learning II

Identifying AD-Sensitive and Cognition-Relevant Imaging Biomarkers via Joint Classification and Regression	115
<i>Hua Wang, Feiping Nie, Heng Huang, Shannon Risacher, Andrew J. Saykin, Li Shen, and ADNI</i>	
Biological Indexes Based Reflectional Asymmetry for Classifying Cutaneous Lesions	124
<i>Zhao Liu, Lyndon Smith, Jiulai Sun, Melvyn Smith, and Robert Warr</i>	
Identification of Paediatric Tuberculosis from Airway Shape Features . . .	133
<i>Benjamin Irving, Pierre Goussard, Robert Gie, Andrew Todd-Pokropek, and Paul Taylor</i>	
Automatic Region-of-Interest Segmentation and Pathology Detection in Magnetically Guided Capsule Endoscopy	141
<i>Philip W. Mewes, Dominik Neumann, Oleg Licegevic, Johannes Simon, Aleksandar Lj. Juloski, and Elli Angelopoulou</i>	
Simultaneous Segmentation and Grading of Hippocampus for Patient Classification with Alzheimer’s Disease	149
<i>Pierrick Coupé, Simon F. Eskildsen, José V. Manjón, Vladimir Fonov, and D. Louis Collins</i>	

Localization of the Lumbar Discs Using Machine Learning and Exact Probabilistic Inference	158
<i>Ayse Betul Oktay and Yusuf Sinan Akgul</i>	
Automatic Contrast Phase Estimation in CT Volumes	166
<i>Michal Sofka, Dijia Wu, Michael Sühling, David Liu, Christian Tietjen, Grzegorz Soza, and S. Kevin Zhou</i>	
3D Shape Analysis for Early Diagnosis of Malignant Lung Nodules	175
<i>Ayman El-Baz, Matthew Nitzken, Ahmed Elnakib, Fahmi Khalifa, Georgy Gimel'farb, Robert Falk, and Mohamed Abou El-Ghar</i>	
Classification of Diffuse Lung Disease Patterns on High-Resolution Computed Tomography by A Bag of Words Approach	183
<i>Rui Xu, Yasushi Hirano, Rie Tachibana, and Shoji Kido</i>	
Discriminative Pathological Context Detection in Thoracic Images Based on Multi-level Inference	191
<i>Yang Song, Weidong Cai, Stefan Eberl, Michael J. Fulham, and Dagan Feng</i>	
X-ray Categorization and Spatial Localization of Chest Pathologies	199
<i>Uri Avni, Hayit Greenspan, and Jacob Goldberger</i>	
Computer-Aided Detection of Ground Glass Nodules in Thoracic CT Images Using Shape, Intensity and Context Features	207
<i>Colin Jacobs, Clara I. Sánchez, Stefan C. Saur, Thorsten Twellmann, Pim A. de Jong, and Bram van Ginneken</i>	
Learning Shape and Texture Characteristics of CT Tree-in-Bud Opacities for CAD Systems	215
<i>Ulas Bagci, Jianhua Yao, Jesus Caban, Anthony F. Suffredini, Tara N. Palmore, and Daniel J. Mollura</i>	
Referenceless Stratification of Parenchymal Lung Abnormalities	223
<i>Sushravya Raghunath, Srinivasan Rajagopalan, Ronald A. Karwoski, Brian J. Bartholmai, and Richard A. Robb</i>	
Lung Texture Classification Using Locally-Oriented Riesz Components	231
<i>Adrien Depeursinge, Antonio Foncubierta-Rodriguez, Dimitri Van de Ville, and Henning Müller</i>	
Computer Aided Diagnosis and Machine Learning III	
Fast Multiple Organ Detection and Localization in Whole-Body MR Dixon Sequences	239
<i>Olivier Pauly, Ben Glocker, Antonio Criminisi, Diana Mateus, Axel Martinez Möller, Stephan Nekolla, and Nassir Navab</i>	

ManiSMC: A New Method Using Manifold Modeling and Sequential Monte Carlo Sampler for Boosting Navigated Bronchoscopy	248
<i>Xiongbiao Luo, Takayuki Kitasaka, and Kensaku Mori</i>	
Characterizing Pathological Deviations from Normality Using Constrained Manifold-Learning	256
<i>Nicolas Duchateau, Mathieu De Craene, Gemma Piella, and Alejandro F. Frangi</i>	
Detecting Outlying Subjects in High-Dimensional Neuroimaging Datasets with Regularized Minimum Covariance Determinant	264
<i>Virgile Fritsch, Gael Varoquaux, Benjamin Thyreau, Jean-Baptiste Poline, and Bertrand Thirion</i>	
Segmenting Images by Combining Selected Atlases on Manifold	272
<i>Yihui Cao, Yuan Yuan, Xuelong Li, Baris Turkbey, Peter L. Choyke, and Pingkun Yan</i>	
Learning Pit Pattern Concepts for Gastroenterological Training	280
<i>Roland Kwitt, Nikhil Rasiwasia, Nuno Vasconcelos, Andreas Uhl, Michael Häfner, and Friedrich Wrba</i>	
Immediate Structured Visual Search for Medical Images	288
<i>Karen Simonyan, Andrew Zisserman, and Antonio Criminisi</i>	
Retrieval Evaluation and Distance Learning from Perceived Similarity between Endomicroscopy Videos	297
<i>Barbara André, Tom Vercauteren, Anna M. Buchner, Michael B. Wallace, and Nicholas Ayache</i>	
Segmentation I	
A New Approach for Tubular Structure Modeling and Segmentation Using Graph-Based Techniques	305
<i>Jack H. Noble and Benoit M. Dawant</i>	
Learning-Based Meta-Algorithm for MRI Brain Extraction	313
<i>Feng Shi, Li Wang, John H. Gilmore, Weili Lin, and Dinggang Shen</i>	
Learning Likelihoods for Labeling (L3): A General Multi-Classifier Segmentation Algorithm	322
<i>Neil I. Weisenfeld and Simon K. Warfield</i>	
Anatomical Labeling of the Anterior Circulation of the Circle of Willis Using Maximum a Posteriori Classification	330
<i>Hroje Bogunović, José María Pozo, Rubén Cárdenes, and Alejandro F. Frangi</i>	

Automatic Multi-organ Segmentation Using Learning-Based Segmentation and Level Set Optimization	338
<i>Timo Kohlberger, Michal Sofka, Jingdan Zhang, Neil Birkbeck, Jens Wetzl, Jens Kaftan, Jérôme Declerck, and S. Kevin Zhou</i>	
Pose-Invariant 3D Proximal Femur Estimation through Bi-planar Image Segmentation with Hierarchical Higher-Order Graph-Based Priors	346
<i>Chaohui Wang, Haithem Boussaid, Loic Simon, Jean-Yves Lazennec, and Nikos Paragios</i>	
Fully Automatic Segmentation of Brain Tumor Images Using Support Vector Machine Classification in Combination with Hierarchical Conditional Random Field Regularization	354
<i>Stefan Bauer, Lutz-P. Nolte, and Mauricio Reyes</i>	
Midbrain Segmentation in Transcranial 3D Ultrasound for Parkinson Diagnosis	362
<i>Seyed-Ahmad Ahmadi, Maximilian Baust, Athanasios Karamalis, Annika Plate, Kai Boetzel, Tassilo Klein, and Nassir Navab</i>	
Segmentation II	
Order Preserving and Shape Prior Constrained Intra-retinal Layer Segmentation in Optical Coherence Tomography	370
<i>Fabian Rathke, Stefan Schmidt, and Christoph Schnörr</i>	
Adaptive Neonate Brain Segmentation	378
<i>M. Jorge Cardoso, Andrew Melbourne, Giles S. Kendall, Marc Modat, Cornelia F. Haggmann, Nicola J. Robertson, Neil Marlow, and Sebastien Ourselin</i>	
Renal Cortex Segmentation Using Optimal Surface Search with Novel Graph Construction	387
<i>Xiuli Li, Xinjian Chen, Jianhua Yao, Xing Zhang, and Jie Tian</i>	
3D Modeling of Coronary Artery Bifurcations from CTA and Conventional Coronary Angiography	395
<i>Rubén Cárdenes, Jose Luis Díez, Ignacio Larrabide, Hrvoje Bogunović, and Alejandro F. Frangi</i>	
Efficient Detection of Native and Bypass Coronary Ostia in Cardiac CT Volumes: Anatomical vs. Pathological Structures	403
<i>Yefeng Zheng, Huseyin Tek, Gareth Funka-Lea, S. Kevin Zhou, Fernando Vega-Higuera, and Dorin Comaniciu</i>	

A <i>Holistic</i> Approach for the Detection of Media-Adventitia Border in IVUS	411
<i>Francesco Ciompi, Oriol Pujol, Carlo Gatta, Xavier Carrillo, Josepa Mauri, and Petia Radeva</i>	
Orientation Histograms as Shape Priors for Left Ventricle Segmentation Using Graph Cuts	420
<i>Dwarikanath Mahapatra and Ying Sun</i>	
Myocardial Segmentation of Late Gadolinium Enhanced MR Images by Propagation of Contours from Cine MR Images	428
<i>Dong Wei, Ying Sun, Ping Chai, Adrian Low, and Sim Heng Ong</i>	
Minimum Average-Cost Path for Real Time 3D Coronary Artery Segmentation of CT Images	436
<i>Ning Zhu and Albert C.S. Chung</i>	
Robust Surface-Based Multi-template Automated Algorithm to Segment Healthy and Pathological Hippocampi	445
<i>Hosung Kim, Tommaso Mansi, Neda Bernasconi, and Andrea Bernasconi</i>	
Model Based 3D Segmentation and OCT Image Undistortion of Percutaneous Implants	454
<i>Oliver Müller, Sabine Donner, Tobias Klinder, Ralf Dragon, Ivonne Bartsch, Frank Witte, Alexander Krüger, Alexander Heisterkamp, and Bodo Rosenhahn</i>	
Optimizing GHT-Based Heart Localization in an Automatic Segmentation Chain	463
<i>Axel Saalbach, Irina Wächter-Stehle, Reinhard Kneser, Sabine Mollus, Jochen Peters, and Jürgen Weese</i>	
Automatic Extraction of 3D Dynamic Left Ventricle Model From 2D Rotational Angiocardialogram	471
<i>Mingqing Chen, Yefeng Zheng, Kerstin Mueller, Christopher Rohkohl, Guenter Lauritsch, Jan Boese, Gareth Funka-Lea, Joachim Hornegger, and Dorin Comaniciu</i>	
Automatic View Planning for Cardiac MRI Acquisition	479
<i>Xiaoguang Lu, Marie-Pierre Jolly, Bogdan Georgescu, Carmel Hayes, Peter Speier, Michaela Schmidt, Xiaoming Bi, Randall Kroecker, Dorin Comaniciu, Peter Kellman, Edgar Mueller, and Jens Guehring</i>	
Multi-part Left Atrium Modeling and Segmentation in C-Arm CT Volumes for Atrial Fibrillation Ablation	487
<i>Yefeng Zheng, Tianzhou Wang, Matthias John, S. Kevin Zhou, Jan Boese, and Dorin Comaniciu</i>	

Accurate and Robust Fully-Automatic QCA: Method and Numerical Validation	496
<i>Antonio Hernández-Vela, Carlo Gatta, Sergio Escalera, Laura Igual, Victoria Martín-Yuste, and Petia Radeva</i>	
Robust Physically-Constrained Modeling of the Mitral Valve and Subvalvular Apparatus	504
<i>Ingmar Voigt, Tommaso Mansi, Razvan Ioan Ionasec, Etienne Assoumou Mengue, Helene Houle, Bogdan Georgescu, Joachim Hornegger, and Dorin Comaniciu</i>	
Regurgitation Quantification Using 3D PISA in Volume Echocardiography	512
<i>Leo Grady, Saurabh Datta, Oliver Kutter, Christophe Duong, Wolfgang Wein, Stephen H. Little, Stephen R. Igo, Shizhen Liu, and Mani Vannan</i>	
Patient-Specific Mitral Leaflet Segmentation from 4D Ultrasound	520
<i>Robert J. Schneider, Neil A. Tenenholtz, Douglas P. Perrin, Gerald R. Marx, Pedro J. del Nido, and Robert D. Howe</i>	
Segmentation III	
Vessel Connectivity Using Murray’s Hypothesis	528
<i>Yifeng Jiang, Zhen W. Zhuang, Albert J. Sinusas, Lawrence H. Staib, and Xenophon Papademetris</i>	
Segmentation of Nerve Bundles and Ganglia in Spine MRI Using Particle Filters	537
<i>Adrian Dalca, Giovanna Danagouliau, Ron Kikinis, Ehud Schmidt, and Polina Golland</i>	
A Multichannel Markov Random Field Approach for Automated Segmentation of Breast Cancer Tumor in DCE-MRI Data Using Kinetic Observation Model	546
<i>Ahmed B. Ashraf, Sara Gavenonis, Dania Daye, Carolyn Mies, Michael Feldman, Mark Rosen, and Despina Kontos</i>	
3-D Graph Cut Segmentation with Riemannian Metrics to Avoid the Shrinking Problem	554
<i>Shouhei Hanaoka, Karl Fritscher, Martin Welk, Mitsutaka Nemoto, Yoshitaka Masutani, Naoto Hayashi, Kuni Ohtomo, and Rainer Schubert</i>	
Adaptive Multi-cluster Fuzzy C-Means Segmentation of Breast Parenchymal Tissue in Digital Mammography	562
<i>Brad Keller, Diane Nathan, Yan Wang, Yuanjie Zheng, James Gee, Emily Conant, and Despina Kontos</i>	

Learning Image Context for Segmentation of Prostate in CT-Guided Radiotherapy	570
<i>Wei Li, Shu Liao, Qianjin Feng, Wufan Chen, and Dinggang Shen</i>	
Joint Thrombus and Vessel Segmentation Using Dynamic Texture Likelihoods and Shape Prior	579
<i>Nicolas Brieu, Martin Groher, Jovana Serbanovic-Canic, Ana Cvejic, Willem Ouwehand, and Nassir Navab</i>	
3D Kidney Segmentation from CT Images Using a Level Set Approach Guided by a Novel Stochastic Speed Function	587
<i>Fahmi Khalifa, Ahmed Elnakib, Garth M. Beache, Georgy Gimelfarb, Mohamed Abou El-Ghar, Rosemary Ouseph, Guela Sokhadze, Samantha Manning, Patrick McClure, and Ayman El-Baz</i>	
Mixture of Segmenters with Discriminative Spatial Regularization and Sparse Weight Selection	595
<i>Ting Chen, Baba C. Vemuri, Anand Rangarajan, and Stephan J. Eisenschenk</i>	
Active Learning for Interactive 3D Image Segmentation	603
<i>Andrew Top, Ghassan Hamarneh, and Rafeef Abugharbieh</i>	
3D Segmentation of Rodent Brain Structures Using Hierarchical Shape Priors and Deformable Models	611
<i>Shaoting Zhang, Junzhou Huang, Mustafa Uzunbas, Tian Shen, Foteini Delis, Xiaolei Huang, Nora Volkow, Panayotis Thanos, and Dimitris N. Metaxas</i>	
Digital Imaging for the Education of Proper Surgical Hand Disinfection	619
<i>Tamás Haidegger, Melinda Nagy, Ákos Lehotsky, and László Szilágyi</i>	
Reinforcement Learning for Context Aware Segmentation	627
<i>Lichao Wang, Robert Merrifield, and Guang-Zhong Yang</i>	
Robust Deformable-Surface-Based Skull-Stripping for Large-Scale Studies	635
<i>Yaping Wang, Jingxin Nie, Pew-Thian Yap, Feng Shi, Lei Guo, and Dinggang Shen</i>	
Confidence-Guided Sequential Label Fusion for Multi-atlas Based Segmentation	643
<i>Daoqiang Zhang, Guorong Wu, Hongjun Jia, and Dinggang Shen</i>	
Probabilistic Multi-shape Segmentation of Knee Extensor and Flexor Muscles	651
<i>Shawn Andrews, Ghassan Hamarneh, Azadeh Yazdanpanah, Bahareh HajGhanbari, and W. Darlene Reid</i>	

Robust Skull Stripping of Clinical Glioblastoma Multiforme Data	659
<i>William Speier, Juan E. Iglesias, Leila El-Kara, Zhuowen Tu, and Corey Arnold</i>	
Multi-stage Learning for Robust Lung Segmentation in Challenging CT Volumes	667
<i>Michal Sofka, Jens Wetzl, Neil Birkbeck, Jingdan Zhang, Timo Kohlberger, Jens Kaftan, Jérôme Declerck, and S. Kevin Zhou</i>	
Author Index	675

Sliding Window and Regression Based Cup Detection in Digital Fundus Images for Glaucoma Diagnosis*

Yanwu Xu¹, Dong Xu¹, Stephen Lin², Jiang Liu³, Jun Cheng³, Carol Y. Cheung⁴,
Tin Aung^{4,5}, and Tien Yin Wong^{4,5}

¹ School of Computer Engineering, Nanyang Technological University, Singapore

² Microsoft Research Asia, P.R. China

³ Institute for Infocomm Research, Agency for Science, Technology and Research, Singapore

⁴ Singapore Eye Research Institute, Singapore

⁵ Department of Ophthalmology, National University of Singapore, Singapore

Abstract. We propose a machine learning framework based on sliding windows for glaucoma diagnosis. In digital fundus photographs, our method automatically localizes the optic cup, which is the primary structural image cue for clinically identifying glaucoma. This localization uses a bundle of sliding windows of different sizes to obtain cup candidates in each disc image, then extracts from each sliding window a new histogram based feature that is learned using a group sparsity constraint. An ϵ -SVR (support vector regression) model based on non-linear radial basis function (RBF) kernels is used to rank each candidate, and final decisions are made with a non-maximal suppression (NMS) method. Tested on the large *ORIGA*^{-light} clinical dataset, the proposed method achieves a 73.2% overlap ratio with manually-labeled ground-truth and a 0.091 absolute cup-to-disc ratio (CDR) error, a simple yet widely used diagnostic measure. The high accuracy of this framework on images from low-cost and widespread digital fundus cameras indicates much promise for developing practical automated/assisted glaucoma diagnosis systems.

1 Introduction

Glaucoma affects about 60 million people [1] and is responsible for approximately 5.2 million cases of blindness (15% of world total) [2]. It unfortunately cannot be cured because the damage to the optic nerve cannot be reversed. Early detection is thus essential for people to seek early treatment and prevent the deterioration of vision [3]. In recent years, much effort has been put into automated/assisted glaucoma diagnosis systems based on computer vision. The design of a glaucoma analysis system depends on the image cues and image modality used.

Among the structural image cues studied for glaucoma diagnosis, those based on the optic disc and cup are of particular importance. The optic disc is located where the ganglion nerve fibers congregate at the retina. The depression inside the optic disc where the fibers leave the retina via the optic nerve head (ONH) is known as the optic cup. The boundaries of the cup and disc structures need to be identified as it facilitates evaluation

* This work is funded by Singapore A*STAR SERC Grant (082 101 0018).

of glaucoma cues such as cup and disc asymmetry and large cup-to-disc ratio (CDR), defined as the ratio of the vertical cup diameter to the vertical disc diameter [4]. The CDR value can be determined by planimetry from color fundus images after the optic disc and cup are outlined manually. Since it is very time consuming and labor intensive to manually annotate the cup and disc for each image, computer vision methods have been proposed to automatically segment the disc and cup in fundus images.

In previous work, researchers have mainly focused on automated segmentation of the optic disc [5], using various techniques such as intensity gradient analysis, Hough transforms, template matching, pixel feature classification, vessel geometry analysis, deformable models and level sets [6] [7]. In this paper, we focus only on the challenging cup detection problem [8] [9], using a large clinical dataset called *ORIGA-light* [10] in which the ground-truth of discs and cups is marked by a team of graders from a hospital. Unlike previous segmentation based algorithms, which classify each pixel as cup or non-cup, our technique identifies a cup as a whole, based on sliding windows and machine learning.

2 Sliding Window Based Cup Detection

In this work, we start with a disc image for cup detection, which may be obtained using methods such as [6]. Different from previous image processing based techniques, a general sliding window based learning framework is proposed for cup localization.

2.1 Sliding Windows

From the suggestion of doctors and graders, in this paper we represent the localized disc by a non-rotated, arbitrary-sized ellipse denoted by its central point (U, V) , corresponding description function $\frac{(x-U)^2}{U^2} + \frac{(y-V)^2}{V^2} = 1$, and rectangular bounding box delimited by $(1, 1)$ and $(2U - 1, 2V - 1)$. With the disc image, we search for the candidate cup by sampling non-rotated ellipses at various aspect ratios represented as $(\mathbf{u}, \mathbf{v}, \mathbf{r}, \mathbf{s})_{N_w \times 4}$, where $(\mathbf{u}, \mathbf{v}, \mathbf{r}, \mathbf{s})$ is the description matrix of all the cup candidates and N_w is the number of cups. For the i^{th} cup candidate denoted as (u_i, v_i, r_i, s_i) , its description function is $\frac{(x-u_i)^2}{r_i^2} + \frac{(y-v_i)^2}{s_i^2} = 1$ and $|r_i| + |u_i| \leq |U|$, $|s_i| + |v_i| \leq |V|$. Cup candidates are generated by sampling values of $(p_i^u, p_i^v, p_i^r, p_i^s)$. In this work, we empirically set $(u_i, v_i, r_i, s_i) = (U \cdot p_i^u, V \cdot p_i^v, U \cdot p_i^r, V \cdot p_i^s)$, where $p_i^u \in [0.75, 1.25]$, $p_i^v \in [0.75, 1.25]$, $p_i^r \in [0.2, 1]$ and $p_i^s \in [0.2, 1]$ with a sampling interval of 0.06. In the detection phase, with this setting, for the input discs with different sizes, $N_w = 6691$ cup candidates from each disc image can be obtained with the same sampling values of $\{(p_i^u, p_i^v, p_i^r, p_i^s) |_{i=1}^{N_w}\}$.

2.2 Feature Representation

Features play an important role in computer vision applications. In this paper, we introduce a new region based color feature for cup detection. Similar to segmentation based approaches, it takes advantage of color differences between cup and disc regions in fundus images. However, it additionally accounts for the elliptical shape of a cup and

the blood vessels that run through it, which often mislead segmentation algorithms. We extract features using the following steps:

1. For a given disc image, the green, blue, hue and saturation channels are computed from the color image. Since the red (RGB model) and value (HSV model) channels differ little between the disc and cup, they are not used in this work. We linearly scale hue and saturation values into $[0,255]$ for consistency with the green and blue color channels. For each color channel, its values are histogrammed by quantization with different bin numbers $\mathcal{B} = \{B_n |_{n=1}^N\}$ such that each bin has an equal (or as equal as possible due to quantization) number of pixels, giving equalized channels. In the experiments, we use $\mathcal{B} = \{3, 4, \dots, 9, 10, 12, 16, \dots, 28, 32\}$.
2. For each color channel and each number of bins $B_n \in \mathcal{B}$, we form three types of features: 1) L1 normalized histogram of the candidate cup region; 2) L1 normalized histogram of the candidate non-cup region within the disc; 3) for each of the B_n bins, the proportion of cup pixels with respect to all the pixels within the disc. Determining the optimal bin numbers in each color channel is non-trivial, so we used multiple bin numbers to generate redundant features and then employ a group sparsity based approach to select the most effective and discriminant features. Finally, each feature is represented as a B_i dimensional vector, and we refer to each type of feature for a given color channel and bin number as a *group*.
3. For a candidate cup in a specific disc, referred to as a ‘‘cup-disc’’ candidate, its original feature \mathbf{f}_i is obtained by concatenating 3 types of features over 4 color channels and multiple bin numbers. In our experimental setting, this leads to a feature dimension of $|\mathbf{f}_i| = \sum_{n=1}^N 3 \times 4 \times B_n = 12 \sum_{n=1}^N B_n = 2208$.

As illustrated in Fig. III after the green channel image is histogrammed into three bins, the first bin (illustrated as black pixels) occupies most of the vessel region, the second bin (grey color) mainly occupies the non-cup region, while the third bin (white color) occupies most of the cup region. Also, it can be observed that the equalized channels are more clear and they facilitate distinguishing different components, since they are relatively insensitive to illumination condition (*e.g.*, see the hue channel). For the cup detection task, it is unclear which color channels to use and how many bins is optimal for a given channel, so we apply statistical learning methods to select features from this large redundant feature representation and use only the selected features for cup localization.

2.3 Feature Selection Based on Group Sparsity Constraint

Identifying and using only the effective elements of the original feature can bring higher precision and speed. For a cup candidate in the training set with an original feature \mathbf{f}_i consisting of g feature groups, we denote its regression value (*i.e.*, the score obtained from its overlap ratio with the clinical ground-truth) as $z_i \in [0, 1]$. We adopt the linear regression model $\omega^T \mathbf{f}_i + \mu$ to obtain the estimated value, where ω is the weighting vector and μ is the bias. We minimize the following objective function:

$$\min_{\omega, \mu} \sum_{i=1}^l \|z_i - \omega^T \mathbf{f}_i - \mu\|^2 + \lambda \sum_{j=1}^g \|\omega_j\|_2 \quad (1)$$

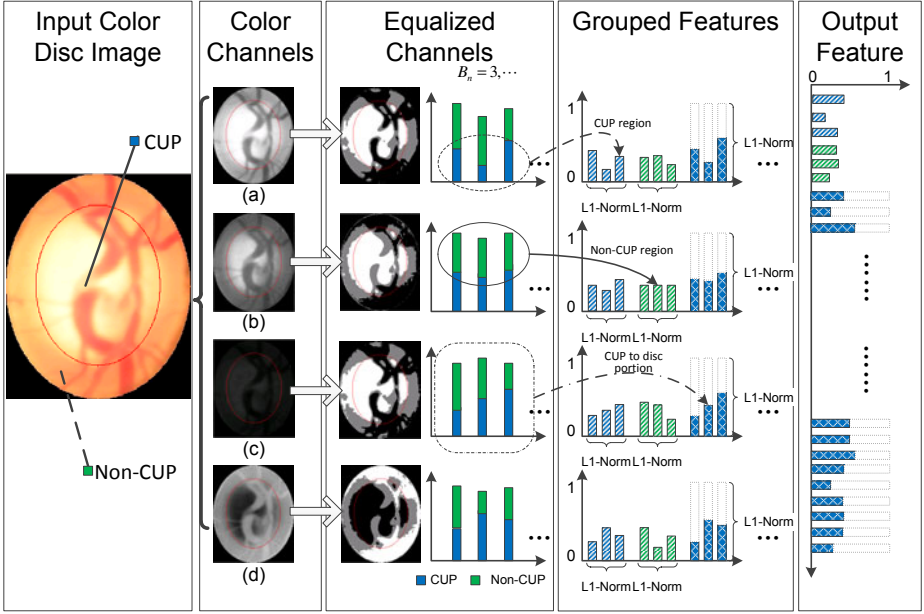


Fig. 1. Grouped feature extraction for cup localization. (a) (b) (c) and (d) represent green, blue, hue and saturation color channels, respectively.

where ω_j is the corresponding weight of the j^{th} feature group, l is the number of training samples and λ is used to control the sparsity of ω . In Eq. (10), the first term represents the regression error and the second term is a $L_{1,2}$ -norm based regularizer to enforce group sparsity. Considering the features are intrinsically organized in groups, we use an $L_{1,2}$ -norm based regularizer to select features from only a sparse set of groups. In the experiments, we use the group-lasso method in [11] to solve Eq. (10).

After ω is obtained, it can be used as a feature selection mask to generate the final features, *i.e.*, the j^{th} group of features is selected when $\|\omega_j\|_2 > 0$. We represent the feature extracted from the i^{th} cup-disc training sample after feature selection as \mathbf{x}_i . The lower dimension of the final feature \mathbf{x}_i leads to faster feature extraction and cup detection in the testing phase when compared with using the original 2208-D feature.

2.4 Non-linear Regression Model

After feature selection, we introduce a kernelized ϵ -SVR to further improve accuracy:

$$\min_{\mathbf{w}, \gamma, \xi, \xi^*} \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^l (\xi_i + \xi_i^*) \quad s.t. \quad \begin{cases} \mathbf{w}^T \phi(\mathbf{x}_i) + b - z_i \leq \epsilon + \xi_i, \\ z_i - \mathbf{w}^T \phi(\mathbf{x}_i) - b \leq \epsilon + \xi_i^*, \\ \xi_i, \xi_i^* \geq 0, i = 1, \dots, l \end{cases} \quad (2)$$

where \mathbf{x}_i is a training sample after feature selection, ξ_i and ξ_i^* are slack variables for ϵ -insensitive loss, C is a regularization parameter, $\mathbf{w}^T \phi(\mathbf{x}_i) + b$ is the non-linear regres-

sion function with \mathbf{w} as the weight vector in the feature space, b as the bias term, and $\phi(\cdot)$ is the non-linear function mapping \mathbf{x}_i from the original space to a higher dimensional space. LibSVM toolbox [12] is used to solve this problem in our implementation.

In the testing phase, the feature \mathbf{x}_i is extracted directly from the i^{th} cup candidate ($i = 1, 2, \dots, N_w$) in the test disc image based on the feature selection mask ω . Then the regression values of all the cup candidates are calculated, denoted as $\gamma = (\gamma_1, \dots, \gamma_i, \dots, \gamma_{N_w})^T$. We sort γ in descending order and obtain the final detection result using the non-maximal suppression (NMS) of the next section.

2.5 Detection Result Fusion with NMS

Various NMS methods have been proposed to reduce redundancy in sliding window based detection. Let us denote the cup candidates as $\mathcal{D} = \{D_1, D_2, \dots, D_{N_w}\}$, where D_i is represented as (u_i, v_i, r_i, s_i) . Note that the cup candidates are sorted according to the regression value γ_i . A detection result can simply be computed as the mean of the top T candidates with the highest regression values, $\overline{D}_T : (\overline{u}_T, \overline{v}_T, \overline{r}_T, \overline{s}_T)$.

Since the top T candidates $D_i|_{i=1}^T$ may not all be of high accuracy, we perform the following steps to handle outliers, similar to majority voting:

1. Initialize a zero matrix $O_{(2U-1) \times (2V-1)}$ of the same size as the disc image.
2. For each cup candidate $D_i|_{i=1}^T$, add a vote for each pixel that lies within D_i .
3. Locate the minimal rectangular bounding box $B_{NMS} : (E_l, E_r, E_t, E_b)$ containing the pixels with no fewer than $\rho \cdot T$ votes, where E_l, E_r, E_t and E_b represent the left, right, top and bottom bounds, respectively, and ρ is a threshold empirically fixed to 0.75 in this work.
4. The final detected cup is represented by the ellipse: $(\frac{E_r+E_l}{2}, \frac{E_t+E_b}{2}, \frac{E_r-E_l+1}{2}, \frac{E_t-E_b+1}{2})$.

3 Experiments

In this section, we describe the evaluation criteria and experimental setting, then analyze the two main steps in our framework, *i.e.*, the group sparsity based feature selection and candidate cup ranking by using RBF based ϵ -SVR, through comparisons of three cup detection methods. The first method (referred to as feature selection+ ϵ -SVR) uses the group sparsity based feature selection method to obtain a low-dimensional feature and then performs RBF based ϵ -SVR to rank the cup candidates. The second method (referred to as feature selection+simple ranking) uses $\omega^T \mathbf{f}_i$ to directly rank the cup candidates after obtaining ω from feature selection. In the third method (referred to as ϵ -SVR), we directly perform RBF based ϵ -SVR ranking using the original feature \mathbf{f}_i without conducting the feature selection process. We also compare our feature selection+ ϵ -SVR approach with level-set based segmentation methods [6] [9].

3.1 Cup Detection Evaluation Criteria

Three evaluation criteria are commonly used for cup detection/segmentation, namely non-overlap ratio (m_1), relative absolute area difference (m_2) [13] and absolute cup-to-disc ratio (CDR) error (δ), defined as:

$$m_1 = 1 - \frac{\text{area}(E_{dt} \cap E_{gt})}{\text{area}(E_{dt} \cup E_{gt})}, m_2 = \frac{|\text{area}(E_{dt}) - \text{area}(E_{gt})|}{\text{area}(E_{gt})}, \delta = \frac{|d_{dt} - d_{gt}|}{R} \quad (3)$$

where E_{dt} denotes a detected cup region, E_{gt} denotes the ground-truth ellipse, d_{dt} is the vertical diameter of the detected cup, d_{gt} is the vertical diameter of the ground-truth cup, $R = 2V - 1$ is the vertical diameter of the disc, and $0 < d_{dt}, d_{gt} \leq R$.

3.2 Experimental Setup

Training samples. The *ORIGA*^{-light} dataset is divided into two sets S_A and S_B , which consist of 150 images and 175 images, respectively. In the training phase, 500 samples including one ground-truth cup and 499 randomly generated cup candidates are obtained for each of the 150 disc images from set S_A . In total, we have 75,000 cup candidates in the training set. The method for generating training cup candidates in the training phase is designed so that the windows of the training cup candidates and those examined in the testing phase have different description parameters (u, v, r, s) . We then use both image sets for testing our algorithm.

Parameter setting for feature selection. For each cup-disc candidate, its original feature \mathbf{f}_i is extracted, and the regression value corresponding to the overlap ratio $(1 - m_1)$ of the cup candidate region and the ground-truth ellipse is also calculated using Eq. (3). Only the ground-truth cup region will have a full score of 1. We solve the problem in Eq. (1) using the group-lasso tool [11] to obtain ω by empirically setting the parameter $\lambda = 0.01$. According to the obtained values of $\|\omega_j\|_2$, 993 of 2208 feature dimensions are selected. Using only 44.97% of the original features leads to significant acceleration in detection speed.

Parameter setting for RBF based ϵ -SVR. The well-known Lib-SVM toolbox [12] is used to train the ϵ -SVR model. We perform cross-validation to determine the optimal parameters by setting the parameters as $C \in \{10^{-3}, 10^{-2}, \dots, 10^2, 10^3\}$, $\epsilon \in \{10^{-3}, 10^{-2}, 10^{-1}\}$, $p \in \{10^{-3}, 10^{-2}\}$, and $g = 2^k \cdot \frac{1}{2\sigma^2}$ with $k \in \{-7, -5, \dots, 5, 7\}$, where p is the convergence threshold in the training phase and σ^2 is the mean of all the Euclidean distances between any two training samples. The samples $\mathbf{x}_i|_{i=1}^l$ are obtained by applying the feature selection mask ω onto the original features $\mathbf{f}_i|_{i=1}^l$. To avoid overlap between the training and testing samples in the cross-validation process, 8000 randomly selected samples and the ground-truth cups from the first 100 images are used for training, while another 6000 randomly selected samples and the ground-truth cups from the remaining 50 images are used for testing. After conducting cross-validation, the optimal parameters were determined to be $C = 10$, $p = 10^{-3}$, $k = -3$ and $\epsilon = 0.001$. With these parameters, all of the 75,000 samples are used to train an ϵ -SVR model for the testing phase.

3.3 Comparison of Three Methods in Our Framework

We compared the three methods to show the effectiveness of each step of our framework. The same cross-validation method is used to determine the optimal parameters of the ϵ -SVR method for a fair comparison. The results are listed in Table 1. From it, we have the following observations:

Table 1. Comparison of three methods in our framework and level-set based methods

Method	Set S_A			Set S_B			$S_A \& S_B$		
	m_1	m_2	δ	m_1	m_2	δ	m_1	m_2	δ
Feature Selection+ϵ-SVR	0.254	0.252	0.081	0.289	0.409	0.106	0.268	0.315	0.091
Feature Sel.+Simple ranking	0.301	0.398	0.115	0.344	0.643	0.143	0.324	0.530	0.130
ϵ -SVR	0.269	0.314	0.101	0.320	0.484	0.128	0.290	0.382	0.112
Level-set [6]	0.458	0.625	0.137	0.552	1.189	0.214	0.495	0.847	0.162
Level-set+Hist-analysis [9]	0.458	0.519	0.119	0.491	0.859	0.159	0.476	0.702	0.140
Relative error reduction to [9]	44.5%	51.5%	31.9%	41.1%	52.4%	33.3%	43.7%	55.1%	35.0%

1. Comparing feature selection+ ϵ -SVR to feature selection+simple ranking shows that the RBF kernel based ϵ -SVR is better than simple ranking using the selected features from our feature selection method. This demonstrates better generalization ability of ϵ -SVR, which is consistent with previous work on image classification.
2. Comparing feature selection+ ϵ -SVR to ϵ -SVR shows that group sparsity based feature selection also improves performance by selecting and using the most effective and discriminant features. Moreover, it accelerates the detection procedure by about 60%. We also observe that the performance improvement from feature selection+ ϵ -SVR over ϵ -SVR is not as large as that from feature selection+ ϵ -SVR over feature selection+simple ranking, possibly because ϵ -SVR also tunes the weight of each feature dimension and thus acts as a kind of feature selection.

3.4 Comparison with Level-Set Based Segmentation [6],[9]

One of the few methods for both cup and disc segmentation is the level-set method of [6], which first identifies the pixels that belong to the cup region, then uses a convex hull method to generate an ellipse. In [9], histogram based analysis of the color pixel intensity together with multiple method fusion are also employed to further improve cup detection accuracy. Table 1 compares our method to these two level-set approaches. Compared with the more advanced approach [9], our method is shown to significantly improve cup localization accuracy in both sets S_A and S_B , and m_1 and CDR error (i.e., δ) are reduced by 43.7% and 35.0%, respectively. We note that all methods obtain better performance on set S_A , possibly because of the data distribution itself. Moreover, it is worth mentioning that the relative CDR error reduction in set S_B is more significant when compared with that in set S_A .

3.5 Detection Speed and Limitations

The experiments were carried on an eight-core 2.67GHz PC with 16GB RAM using the Matlab Parallel Computing Toolbox. In our approach, the extraction of the original feature takes about 6 minutes per image, while feature selection reduces the time

¹ We did not compare with the Haar+Adaboost method for general object detection, because the cup detection task was formulated as a regression problem, not a binary classification problem, and the Haar feature is not suitable for objects with varying aspect ratios.

cost by about 60%. The RBF based ϵ -SVR takes about 1 minute per image. The NMS takes about 0.2 minutes per image on average. The main time cost is for feature extraction, and the proposed sparsity based feature selection greatly accelerates the detection speed. In addition, from our observations the proposed method does not handle large cups as effectively, because NMS suppresses the rim of the cup.

4 Conclusion

We proposed a sliding window based learning framework with a newly developed feature for cup detection in glaucoma diagnosis. Tested on a large clinical dataset with three evaluation criteria, it achieves a 26.8% non-overlap ratio (m_1) with manually-labeled ground-truth, a 31.5% relative absolute area difference (m_2) and a 0.091 absolute CDR error (δ). In future work, we plan to elevate performance using new features or by introducing domain-specific knowledge on this problem.

References

1. Wong, T., Loon, S., Saw, S.: The epidemiology of age related eye diseases in asia. *British Journal of Ophthalmology* 90(4), 506–511 (2006)
2. Thylefors, B., Negrel, A.: The global impact of glaucoma. *Bull. World Health Organ.* 72(3), 323–326 (2006)
3. Michelson, G., Wartges, S., Hornegger, J., Lausen, B.: The papilla as screening parameter for early diagnosis of glaucoma. *Dtsch. Arztebl. Int.* 105(34-35), 583–589 (2008)
4. Jonas, J., Budde, W., Panda-Jonas, S.: Ophthalmoscopic evaluation of the optic nerve head. *Survey of Ophthalmology* 43, 293–320 (1999)
5. Abramoff, M., Alward, W., Greenlee, E., Shuba, L., Kim, C., Fingert, J., Kwon, Y.: Automated segmentation of the optic disc from stereo color photographs using physiologically plausible features. *Investigative Ophthalmology and Vis. Sci.* 48(4), 1665–1673 (2007)
6. Liu, J., Wong, D., Lim, J., Li, H., Tan, N., Zhang, Z., Wong, T., Lavanya, R.: ARGALI: an automatic cup-to-disc ratio measurement system for glaucoma analysis using level-set image processing. In: *Int. Conf. Biomed. Eng.* (2008)
7. Li, C., Xu, C., Gui, C., Fox, M.: Level set evolution without re-initialization: A new variational formulation. In: *IEEE Conf. Comput. Vis. Patt. Rec.*, pp. 430–436 (2005)
8. Merickel, M., Wu, X., Sonka, M., Abramoff, M.: Optimal segmentation of the optic nerve head from stereo retinal images. In: *Medical Imaging: Physiology, Function, and Structure from Medical Images*, vol. 6143 (2006)
9. Wong, D., Liu, J., Lim, J., Tan, N., Zhang, Z., Lu, S., Li, H., Teo, M., Chan, K., Wong, T.: Intelligent fusion of cup-to-disc ratio determination methods for glaucoma detection in ARGALI. In: *Int. Conf. Engin. in Med. and Biol. Soc.*, pp. 5777–5780 (2009)
10. Zhang, Z., Yin, F., Liu, J., Wong, D., Tan, N., Lee, B., Cheng, J., Wong, T.: Origa-light: An online retinal fundus image database for glaucoma analysis and research. In: *Int. Conf. IEEE Engin. in Med. and Biol. Soc.*, vol. 2010, pp. 3065–3068 (2010)
11. Jacob, L., Obozinski, G., Vert, J.P.: Group lasso with overlap and graph lasso. In: *Int. Conf. Machine Learning. ACM, New York* (2009)
12. Chang, C., Lin, C.: LIBSVM: a library for support vector machines. *ACM Trans. on Intel. Sys. and Tech.* 2(3), 27:1–27:27 (2011)
13. Deng, X., Du, G.: Liver tumor segmentation. In: *3D Segmentation in the Clinic: A Grand Challenge II Workshop at 10th MICCAI* (2007)

Learning from Only Positive and Unlabeled Data to Detect Lesions in Vascular CT Images

Maria A. Zuluaga^{1,2}, Don Hush³, Edgar J.F. Delgado Leyton^{1,2},
Marcela Hernández Hoyos², and Maciej Orkisz¹

¹ CREATIS; Université de Lyon; Université Lyon 1; INSA-Lyon; CNRS UMR5220;
INSERM U1044; F-69621 Villeurbanne, France

² Grupo Imagine, Grupo de Ingeniería Biomédica,
Universidad de los Andes, Bogotá, Colombia

³ Los Alamos National Laboratory, Los Alamos, NM 87545, USA

Abstract. Detecting vascular lesions is an important task in the diagnosis and follow-up of the coronary heart disease. While most existing solutions tackle calcified and non-calcified plaques separately, we present a new algorithm capable of detecting both types of lesions in CT images. It builds up on a semi-supervised classification framework, in which the training set is made of both unlabeled data and a small amount of data labeled as normal. Our method takes advantage of the arrival of newly acquired data to re-train the classifier and improve its performance. We present results on synthetic data and on datasets from 15 patients. With a small amount of labeled training data our method achieved a 89.8% true positive rate, which is comparable to state-of-the-art supervised methods, and the performance can improve after additional iterations.

1 Introduction

As the evaluation of coronary lesions is challenging and tedious, and acquiring moderate expertise in coronary CT angiography (CTA) may take more than one year [1], a variety of methods has been proposed to perform this detection automatically. Most of the algorithms have been directed towards abnormality modeling, *i.e.* identifying the particularities of lesions. As the latter are heterogeneous by nature, and obtaining a model that copes with all possible abnormalities is difficult, most approaches have tackled the segmentation of one type of lesion: calcified plaques [2] or soft plaques [3]. However, these methods often rely on different image acquisition techniques for each task (*e.g.* non-enhanced CT for calcium quantification, contrast-enhanced CT for lumen segmentation and Dual-Source CT for soft plaque locating), which makes it difficult to combine them, in order to simultaneously tackle the automated detection of both types of plaques. Recently, a novel family of methods has made use of machine learning techniques to simultaneously detect the different forms of lesions [4]. Since such methods use supervised classification schemes, the set of examples used for training has to be highly reliable. Unfortunately, it is very expensive to collect labeled data that are accurate, as well as representative of all types of lesions.

A first attempt to use an unsupervised scheme has been proposed in [5]. However, it failed to distinguish bifurcations from actual lesions.

To overcome these shortcomings, here we propose to use a semi-supervised classification scheme that focuses in the healthy vessel sections. It permits both calcified and non-calcified lesions to be identified as the complement of the healthy sections. To do so, we introduced a classification algorithm belonging to a family named Learning from only Positive and Unlabeled data (LPU). Only a relatively small amount of healthy sections must be provided at the learning stage of this algorithm. Its use to identify vascular lesions is novel to the best of our knowledge. We also proposed a strategy that exploits new data that daily arrive in a clinical environment, in order to refine the learning and thus improve the classification performance. Our software has been made publicly available¹.

2 Method

In the context of lesion identification, classification methods try to differentiate between two main classes: the healthy and the diseased one. However, while the appearance of healthy vascular sections does not vary much, the appearance of the diseased ones may show a large variability. In other words, the *healthy* class is likely to form a dense cluster in the feature space, while the *diseased* class is represented by sparse points rather than by a cluster. Moreover, we consider that obtaining reliable labels of only healthy vessel sections is an easier task than obtaining representative examples of all types of diseased vessels. Based on these two statements, we addressed the vascular lesion identification through an LPU framework using support vector machines (SVM) [6]. Such an algorithm can be applied to the problems where the training input data is made up of labeled samples (the healthy class) and a large amount of unlabeled samples coming from the mixture (healthy and diseased samples).

2.1 Learning from Only Positive and Unlabeled Samples

Let $Q = (q_1, \dots, q_k), q_i \in \mathcal{X}$ be a collection of samples generated according to a probability distribution P_q and $X = (x_1, \dots, x_n), x_i \in \mathcal{X}$ a set generated according to a probability distribution P_x . The goal is to identify the elements of X that are similar to the elements of Q . It is assumed that elements in Q belong to a particular class (our healthy class), and that elements belonging to X are actually formed by two different distributions. Some are generated by P_q (healthy samples) and the rest (the diseased samples) by another process P_{other} . Therefore, P_x is a mixture distribution:

$$P_x = \beta P_q + (1 - \beta) P_{other} \quad \text{where } 0 < \beta < 1. \quad (1)$$

Porter *et al.* [6] have defined a relative content density function p that can be used as a similarity measure to quantify the relative concentration of P_q with respect

¹ <http://www.creatis.insa-lyon.fr/software/public/DLalgorithms/>

Table 1. The LPU algorithm

Input	: Q set- Samples of healthy vessel cross-sections
Input	: X set- Samples of unlabeled vessel cross-sections
1	: Assign y=1 labels to Q
2	: Assign y=-1 labels to X
3	: Assign the y=1 class weight of a SVM to 2β
4	: Assign the y=-1 class weight of a SVM to $1-2\beta$
5	: Train the SVM Select f that minimizes $\mathcal{R}(f)$
6	: Classify samples in X using the trained model M
Output	: Classified X samples

to P_x , and which is given by the Radon-Nikodym derivative $p = dP_q/dP_x$. Given a threshold ρ , the set $\{x \in X : p(x) > \rho\}$ contains samples that are more likely to be generated by P_q than by P_{other} .

The latter problem can be solved as a density level detection problem, where a function f is constructed so that it approximates the set $\{p > \rho\}$ by means of the set $\{f > 0\}$. Steinwart *et al.* [7] have proposed a risk function \mathcal{R} that can be estimated from sample data, and serves as a criterion to assess the quality of the approximation of $\{p > \rho\}$ by $\{f > 0\}$:

$$\mathcal{R}(f) = \frac{2\beta}{1+2\beta} \frac{1}{|Q|} \sum_{x \in Q} I(f(x) \leq 0) + \frac{1}{1+2\beta} \frac{1}{|X|} \sum_{x \in X} I(f(x) > 0), \quad (2)$$

where $\rho = \frac{1}{2\beta}$, $I(\cdot) = 1$ if the argument is true and 0 otherwise.

The main advantage of defining such a risk function \mathcal{R} is that it allows us to use a SVM in order to choose a function f that minimizes \mathcal{R} [7]. For this purpose, a surrogate problem is constructed by automatically assigning labels y to the available data. The required steps to build the surrogate problem and solve the LPU algorithm are outlined in Table 1.

Let us note however, that the LPU formulation is theoretically valid in the infinite sample limit, *i.e.* $|X| = \infty$. Additionally, the problem we tackle is highly unbalanced, *i.e.* the cardinality of one class is much larger than that of the other one (in our case, this corresponds to a true ρ value close to 1). With unbalanced classes most learning methods tend to favor a response that assigns all samples to one class, which can worsen when combined with finite sample effects. Since infinite samples cannot be achieved, it is at least desired that $|X| \gg |Q|$.

2.2 LPU in a Clinical Environment

In the LPU formulation unlabeled data makes part of the training set. This key difference w.r.t. supervised classification approaches is an advantage in two aspects. First, the training set can be easily augmented since no labels are required for the X set. Second, it is desirable and feasible to increase the training set without great risk of overfitting.

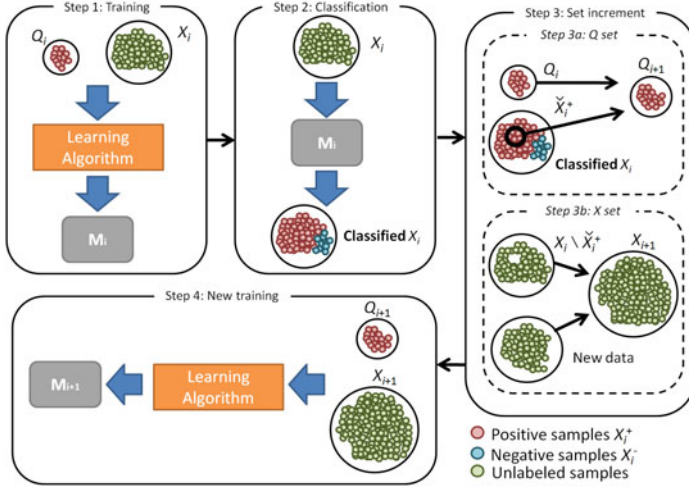


Fig. 1. LPU in a clinical environment. Positive samples are those identified as healthy (either by the algorithm or an expert), negative ones are those classified as diseased, and unlabeled ones are samples that belong to the X set before classification.

Based on this key feature, we proposed to progressively increase the two sets that make up the training data set: the addition of samples to Q aimed at normality description improvement, whereas the increase of X kept the relation $|X| \gg |Q|$ and reduced the finite sample effect. Consequently, we suggested to periodically re-train the model M after new data incorporation, instead of keeping it static. Our proposal suits the clinical data workflow where new unlabeled data arrive daily. Moreover, no additional labeling is required to exploit the arriving data. The algorithm (Fig. 1) starts with an initial pair of sets, unlabeled data X_1 and (manually) labeled data Q_1 , then it iterates as follows:

1. **Training:** A pair of sets Q_i and X_i is used to train a model M_i .
2. **Classification:** The model M_i is used to label the data from X_i . A clinician validates the labels corresponding to actual lesions (subset denoted X_i^-).
3. **Set increase:** This step is performed for both Q and X sets.
 - (a) Among the samples that the clinician did not consider as lesions (subset denoted X_i^+) a sub-subset \tilde{X}_i^+ is randomly removed from the set X_i and combined with the set Q_i to build up the increased set $Q_{i+1} = Q_i \cup \tilde{X}_i^+$.
 - (b) At the arrival of newly acquired data, this is combined with the subset $X_i \setminus \tilde{X}_i^+$ to make up the increased set X_{i+1} .
4. **Loop:** The algorithm jumps to the step 1 in order to obtain a new refined model M_{i+1} by using the increased sets Q_{i+1} and X_{i+1} for training. For this purpose, the labels previously assigned to X_i are considered as unavailable.

2.3 Features

To select the features that fit our problem we followed some of the guidelines in [8]. We used a combination of metrics that are able to capture both the circular

shape with radially decreasing intensity profiles of healthy vessel cross-sections, and the divergence from such typical patterns. However, we used a different selection strategy based on the empirical risk \mathcal{R} minimization. Four metrics calculated in cross-sections orthogonal to the vessel centerline (Concentric rings [5], Core, Hessian eigenvalues and Flux [9]) were kept.

3 Experiments and Results

The LPU algorithm was solved using the LIBSVM software [10] with a Gaussian radial basis function (RBF) kernel. The SVM regularization term λ and the σ^2 parameter of the RBF kernel were optimized through a grid search. Different $\beta \in [0.1, 0.2, \dots, 0.80, 0.81, 0.82, \dots, 0.98, 0.99]$ values were tested. The selection of the optimal β value was included in the optimization process. For every $\{\lambda, \sigma, \beta\}$ combination, the empirical risk \mathcal{R} associated with f , was calculated according to Equation 2. The learned decision function f , minimizing the empirical risk, was applied on the X set to evaluate the performance of the method.

We first evaluated our method on synthetic images with known ground truth. We then applied it to 3D cardiac CT data sets, where lesions had been annotated by an expert for the purpose of evaluation only. In our experiments, lesion detection was evaluated using three measures: true positive rate $TPR = \frac{TP}{TP+FN}$, true negative rate $TNR = \frac{TN}{TN+FP}$ and balanced error rate $BER = 1 - \frac{TPR+TPN}{2}$. Here TP denotes the number of correctly classified diseased cross-sections, TN the number of correctly classified healthy cross-sections, while FP and FN respectively are the numbers of falsely classified healthy and diseased cross-section. We preferred the use of BER instead of the commonly used accuracy, since the latter is not very meaningful in highly unbalanced problems as the one we tackle.

3.1 Synthetic Data

We first evaluated the performance of our method on 70 artificially generated volumes containing a variety of cases typically encountered in vascular analysis. Phantoms were created using the typical Hounsfield Unit values that are found in CT images for blood, background and plaque components, as well as the typical image dimensions and voxel size. Gaussian noise was also added, resulting in a contrast-to-noise-ratio value of 10.

To demonstrate the effectiveness of the iterative increase of the training set, we simulated five iterations. The cardinality ($|Q|, |X|$) at each iteration was as follows: (500, 15000), (1000, 19000), (1200, 22000), (1300, 22000), (1600, 23000). A sixth case aimed at illustrating the situation, where the condition $|X| \gg |Q|$ is broken, with $(|Q|, |X|) = (1600, 2000)$.

The LPU algorithm had a good performance in terms of TPR, which is desirable for disease detection (Fig. 2). As more samples were included in the sets, the TNR improved. Since the TPR remained unchanged, the total error (BER) decreased. The 6-th case confirmed the expected poor performance of LPU when the condition $|X| \gg |Q|$ is broken.

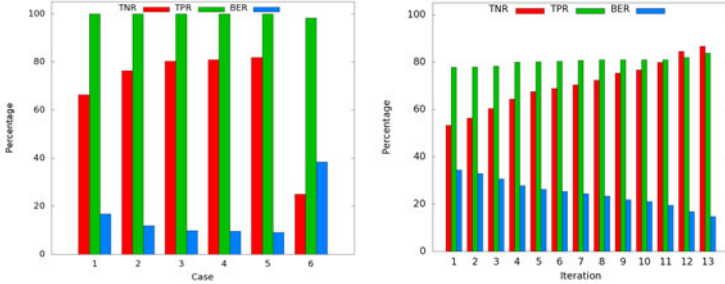


Fig. 2. Evolution of TNR, TPR and BER as a consequence of the modification of Q and X datasets. *Left*, 5 iterations in synthetic data, and a 6-th case, where $|X| \gg |Q|$ does not hold. *Right*, 13 iterations using patient data. Final TNR, TPR and BER values were 86.7%, 83.8% and 14.2%, respectively.

3.2 Patient Data

Fifteen cardiac CT datasets, with centerlines available in a total of 53 arteries, were obtained from two different sources: Hôpital Louis Pradel (Bron, France) and Rotterdam Coronary Artery Algorithm Evaluation Framework [11], the latter containing data acquired at the Erasmus Medical Center (Rotterdam, The Netherlands). Cross-sections orthogonal to the centerlines were calculated, from which the required features were extracted. Additionally, each cross-section was labeled as normal or abnormal by an observer. These annotations were used to evaluate the performance of the classifier and a small percentage of annotated normal cross-sections were used to build up the initial set Q_1 . At the first iteration, we used normal data from a healthy subject ($|Q_1| = 426$) and unlabeled cross-sections ($|X_1| = 1148$) from two patients. Due to a limited amount of annotated data available for evaluation, only one patient was added at every new iteration. This allows us to demonstrate the behavior of our method in real data, although in clinical practice several patients may arrive daily. At the final iteration, the cardinalities of the sets were: $|Q_{13}| = 886$, $|X_{13}| = 12180$.

The evolution of the TNR, TPR and BER as a function of the iterations is presented in Figure 2. Similarly to the synthetic data, the error decreased as new samples were added to Q and X . LPU showed a tendency to overestimate the lesions, possibly because the number of labeled samples $|Q| = 886$ used for training was not yet sufficiently representative of all normal configurations. However, despite a small number of training data, the TNR increased from 0.53 to 0.86 as new samples were added, *i.e.* the false alarm rate ($1 - \text{TNR}$) decreased from 0.47 to 0.14. From the tendency of TPR, TNR and BER (Fig. 2), we believe that lesion overestimation can be reduced by using LPU with an increased training set. Figure 3 shows examples of classified cross-sections in several coronary arteries.

The same definition of TP has been used in [5], so a direct comparison of the results can be done. In that work, a $\text{TPR} = 0.860$ and $\text{TNR} = 0.812$ have been reported. From this information we have computed the $\text{BER} = 0.164$. While, our new proposal gives a lower $\text{TPR} = 0.838$, it has a higher $\text{TNR} = 0.867$ that can be

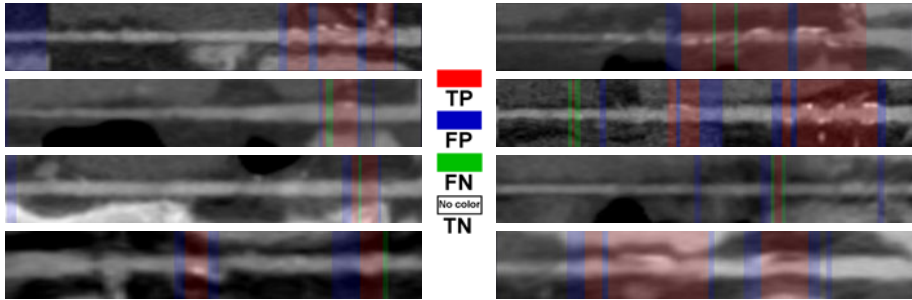


Fig. 3. Lesion detection results. Color-coded labels on a stretched curved planar reformatted view of different coronary arteries.

explained by the fact that our novel method does not misclassify bifurcations, contrary to the one in [5]. Our lower BER=0.142 confirms that our overall performance is better. Another related work [4] lacks a clear definition of the evaluation measurement unit, making direct comparison uneasy. Their evaluation has been performed in terms of detected lesions, *i.e.* a lesion is correctly detected (TP) if at least one cross-section within the extent of the lesion is detected as diseased. The available information permitted us to compute their TPR=0.890. Using the latter definition of TP, our approach gave a slightly better TPR=0.898.

4 Conclusions

We proposed a new semi-supervised algorithm to detect coronary artery lesions in CTA images. The method can achieve a high detection rate (TPR) even when a small amount of labeled training data is available. Its false alert rate (1-TNR) can be substantially reduced by increasing the training set, as new data arrive and new normal samples are validated. The improvement is guaranteed since adding new samples reduces the finite sample problem. However, recovery is not guaranteed if the user erroneously validates misclassified samples. The computational time increases with the size of the sets Q and X . After a number of iterations, the classification error may not significantly decrease despite the inclusion of additional samples. Therefore, the optimal trade-off between computational time and capacity of improvement is to be carefully evaluated. Let us note that the LPU formalism uses only one type of labels (here 'normal') to train the classifier. It is not straightforward to use the diseased samples correctly labeled in previous iterations, in order to train the model in the next iterations. Future work should attempt to overcome this limitation.

Acknowledgments. This work was partly funded by the projects THROMBUS (European Commission's 7th Framework Program), ECOS Nord C11S01, Colciencias 1204-519-28996 and UniAndes Interfacultades 06-2010.

References

1. Pugliese, F., Hunink, M.G.M., Gruszczynska, K., Alberghina, F., Malagó, R., van Pelt, N., Mollet, N.R., Cademartiri, F., Weustink, A.C., Meijboom, W.B., Witteman, C.L.M., de Feyter, P.J., Krestin, G.P.: Learning curve for coronary CT angiography: what constitutes sufficient training? *Radiology* 251, 359–368 (2009)
2. Saur, S.C., Alkadhi, H., Desbiolles, L., Székely, G., Cattin, P.C.: Automatic detection of calcified coronary plaques in computed tomography data sets. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) *MICCAI 2008, Part I. LNCS*, vol. 5241, pp. 170–177. Springer, Heidelberg (2008)
3. Makrogiannis, S., Bhotika, R., Miller, J.V., Skinner Jr., J., Vass, M.: Nonparametric intensity priors for level set segmentation of low contrast structures. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) *MICCAI 2009, Part I. LNCS*, vol. 5761, pp. 239–246. Springer, Heidelberg (2009)
4. Tessmann, M., Vega-Higuera, F., Fritz, D., Scheuring, M., Greiner, G.: Multi-scale feature extraction for learning-based classification of coronary artery stenosis. In: *Proc. SPIE Med. Imaging*, vol. 7260 (2009), doi:10.1117/12.811639
5. Zuluaga, M.A., Magnin, I.E., Hernández Hoyos, M., Delgado Leyton, E.J.F., Lozano, F., Orkisz, M.: Automatic detection of abnormal vascular cross-sections based on Density Level Detection and Support Vector Machines. *Int. J. Computer Assisted Radiol. Surg.* 6, 163–174 (2011)
6. Porter, R., Ruggiero, C., Hush, D.: Density-based similarity measures for content based search. In: *43rd Asilomar Conf. Signals Systems Computers*, pp. 390–394 (2009)
7. Steinwart, I., Hush, D., Scovel, C.: A classification framework for anomaly detection. *J. Mach. Learning Research* 6, 211–232 (2005)
8. Zuluaga, M.A., Delgado Leyton, E.J.F., Hernández Hoyos, M., Orkisz, M.: Feature selection for SVM-based vascular anomaly detection. In: Menze, B., Langs, G., Tu, Z., Criminisi, A. (eds.) *Medical Computer Vision. Recognition Techniques and Applications in Medical Imaging. LNCS*, vol. 6533, pp. 141–152. Springer, Heidelberg (2010)
9. Lesage, D., Angelini, E.D., Bloch, I., Funka-Lea, G.: A review of 3D vessel lumen segmentation techniques: models, features and extraction schemes. *MedIA* 13(6), 819–845 (2009)
10. Chang, C.-C., Lin, C.-J.: LIBSVM: a library for support vector machines (2004), <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>
11. Schaap, M., Metz, C., van Walsum, T., van der Giessen, A.G., Weustink, A.C., Mollet, N.R., Bauer, C., Bogunovic, H., Castro, C., Deng, X., Dikici, E., O'Donnell, T., Frenay, M., Friman, O., Hernández Hoyos, M., Kitslaar, P.H., Krissian, K., Kühnel, C., Luengo-Oroz, M.A., Orkisz, M., Smedby, Ö., Styner, M., Szymczak, A., Tek, H., Wang, C., Warfield, S.K., Zambal, S., Zhang, Y., Krestin, G.P., Niessen, W.J.: Standardized evaluation methodology and reference database for evaluating coronary artery centerline extraction algorithms. *MedIA* 13(5), 704–715 (2009)

Regularized Tensor Factorization for Multi-Modality Medical Image Classification

Nematollah Batmanghelich, Aoyan Dong, Ben Taskar, and Christos Davatzikos

Section for Biomedical Image Analysis,
Suite 380, 3600 Market St., 19104 Philadelphia, US
{batmangh@seas, aoyan.dong@uphs, taskar@cis, christos@rad}.upenn.edu
<http://www.rad.upenn.edu/sbia>

Abstract. This paper presents a general discriminative dimensionality reduction framework for multi-modal image-based classification in medical imaging datasets. The major goal is to use all modalities simultaneously to transform very high dimensional image to a lower dimensional representation in a discriminative way. In addition to being discriminative, the proposed approach has the advantage of being clinically interpretable. We propose a framework based on regularized tensor decomposition. We show that different variants of tensor factorization imply various hypothesis about data. Inspired by the idea of multi-view dimensionality reduction in machine learning community, two different kinds of tensor decomposition and their implications are presented. We have validated our method on a multi-modal longitudinal brain imaging study. We compared this method with a publically available classification software based on SVM that has shown state-of-the-art classification rate in number of publications.

Keywords: Tensor factorization, Multi-view Learning, Multi-Modality, Optimization, Basis Learning, Classification.

1 Introduction

Recently, various structural (*e.g.* MRI, DTI, *etc.*) and functional (*e.g.* PET, resting state fMRI, *etc.*) imaging modalities have been utilized to develop new biomarkers for diagnosis. Multiple image modalities can provide a rich multi-parametric signature that can be used to design more sensitive biomarkers [12], [10], [14]. For example, while structural MR images provide sensitive measurements for detection of atrophy in brain regions [8], FDG-PET¹ can quantify reduction of glucose metabolism in parietal lobes, the posterior cingulate, and other brain regions [5]; combination of both modalities can be very instrumental in early diagnosis of Alzheimer's disease [7].

An immediate solution to exploit multiple modalities is to concatenate all image modalities into a long vector, but learning a classifier that generalizes well in such a high dimensional space is even harder than in the uni-modality case

¹ fluorodeoxyglucose positron emission tomography.

because multi-modality datasets tend to be small. Therefore, dimensionality reduction plays an even more important role here. Most existing studies extract features from a few predefined areas [12]. Zhang *et al.* [14] suggested extracting features from a few pre-defined regions of interest (ROIs) and combining them into one kernel that then input to a kernel-SVM classifier. However, predefined regions might not be optimal for diagnosis on the individual level, *i.e.* classification of subjects into normal and abnormal groups. Ideally, whole image (*e.g.* brain scan) should be viewed as a large dimensional observation and relevant regions to the target variable of interest (class labels, here) should be derived from such high dimensional observation. High-dimensional pattern classification methods have been proposed for morphological analysis [6], [9] which aim to capture multivariate nonlinear relationships in the data. A critical step underlying the success of such methods is effective feature extraction and selection, *i.e.* dimensionality reduction. Batmanghelich *et al.* [2] used a constrained matrix factorization framework for dimensionality reduction while simultaneously being discriminative and representative; however, that method only works for uni-modality cases. In this paper, we propose a method inspired by the *multi-view* setting in the machine learning community [11], [1]. In the multi-view setting, there are views (sometimes in a rather abstract sense) of the data which co-occur, and there is a target variable of interest (class labels, here). The goal is to implicitly learn the target via the relationship between different views [11]. Our approach extends [2] to tensor factorization framework to handle the multi-modality case, but our formulation and optimization method is substantially different.

One could concatenate all image modalities of a subject into long columns of a matrix and simply apply [2] or a similar method. However, the advantage of extending a regularized matrix factorization to a tensor factorization framework is that because of the structure of a tensor, various factorizations can be proposed, each of which imply different hypotheses about the data. In this paper, we introduce two factorizations and explain their connotations. We derive the factorization by solving a large scale optimization problem.

2 General Framework

The novel method proposed in this paper is based on an extension of a previously presented framework for uni-modality [3], which we briefly present here for perspective. Similar to [2], the proposed method reduces the dimensionality in a discriminative way while preserving the semantics of images; hence it is clinically interpretable and produces good classification accuracy. We use regularized matrix factorization formalism for dimensionality reduction. Regularized matrix factorization decomposes a matrix into two or more matrices such that the decomposition describes the matrix as accurately as possible. Such a decomposition could be subjected to some constraints or priors. Let us assume columns of $\mathbf{X} = [\mathbf{x}_1 \cdots \mathbf{x}_n \cdots \mathbf{x}_N]$ represent observations (*i.e.* sample images that are vectorized), and $\mathbf{B} \in \mathfrak{R}^{D \times K}$ and $\mathbf{C} \in \mathfrak{R}^{K \times N}$ decompose the matrix such that

$\mathbf{X} \approx \mathbf{BC}$. K is the number of basis vectors, D is the number of voxels in images and N is the number of samples. The columns of matrix \mathbf{B} (called \mathbf{b}_k) can then be viewed as basis vectors and the n^{th} column of \mathbf{C} (called \mathbf{c}_n) contains corresponding loading coefficients or weights of the basis vectors for the n^{th} observation. The columns $\mathbf{b}_k \in \mathcal{B}$ and $\mathbf{c}_n \in \mathcal{C}$ are subjected to some constraints which define the feasible sets \mathcal{B} and \mathcal{C} . We use variable $y_n \in \{-1(\text{abnormal}), 1(\text{healthy})\}$ to denote labels of the subjects.

An optimal basis vector (\mathbf{b}_k) operates as a region selector; therefore its entries (b_{jk}) must be either *on* (i.e. 1) or *off* (i.e. 0) (i.e. $b_{jk} \in \{0, 1\}$). Since optimizing integer values is computationally expensive, particularly for the large dimensionality characteristic of medical images, we relax this constraint to $0 \leq b_{jk} \leq 1$ which can be encoded mathematically by a combination of ℓ_∞ norm and non-negativity ($\mathbf{b} \geq 0$). Assuming that only certain structures of an anatomy are affected (e.g. atrophy of hippocampus in Alzheimer’s disease), we can impose sparsity on the basis vectors which also make them more interpretable. The sparsity constraint can be enforced by an inequality constraint over the ℓ_1 norm of the basis vectors. These two properties constitute the feasible set for the basis vectors (\mathcal{B}) as follows:

$$\mathcal{B} := \{\mathbf{b} \in \mathfrak{R}^D : \mathbf{b} \geq \mathbf{0}, \|\mathbf{b}\|_\infty \leq 1, \|\mathbf{b}\|_1 \leq \lambda_3\}$$

where the ratio of λ_3/D encodes the ratio of sparsity of the basis vectors.

For the feasible set of coefficients (\mathcal{C}), we only assume non-negativity (i.e. $\mathcal{C} := \{\mathbf{c} : \mathbf{c} \geq \mathbf{0}\}$) because our images are non-negative but this is relaxable based on the properties of a problem.

In order to find optimal \mathbf{B} and \mathbf{C} matrices, we define the following constrained optimization problem:

$$\begin{aligned} \min_{\mathbf{B}, \mathbf{C}, \mathbf{w} \in \mathfrak{R}^K} \quad & \lambda_1 \mathcal{D}(\mathbf{X}; \mathbf{BC}) + \lambda_2 \sum_{n=1}^N \ell(y_n; f(\mathbf{x}_n; \mathbf{B}, \mathbf{w})) + \|\mathbf{w}\|_2 \\ \text{subject to:} \quad & f(\mathbf{x}_n; \mathbf{B}, \mathbf{w}) = \langle \mathbf{B}^T \mathbf{x}_n, \mathbf{w} \rangle \\ & \mathbf{b}_k \in \mathcal{B}, \quad \mathbf{c}_i \in \mathcal{C} \end{aligned} \tag{1}$$

The cost function of the optimization problem consists of two terms: 1) The *generative* term ($\mathcal{D}(\cdot; \cdot)$) encourages the decomposition, \mathbf{BC} , to be close to the data matrix (\mathbf{X}); both labeled and unlabeled data contribute to this term. 2) The *discriminative* term ($\ell(y_n; f(\mathbf{x}_n, \mathbf{B}, \mathbf{w}))$) is a *loss* function that encourages a classifier $f(\cdot)$ to produce class labels that are consistent with available labels (\mathbf{y}). The classifier parametrized by \mathbf{w} projects each image (\mathbf{x}_n) on the basis vectors to produce new features ($\mathbf{v}_n = \mathbf{B}^T \mathbf{x}_n$) and produce a label. We use a linear classifier, hence $f(\mathbf{x}_n, \mathbf{B}, \mathbf{w}) = \langle \mathbf{B}^T \mathbf{x}_n, \mathbf{w} \rangle$. In this paper, we set $\mathcal{D}(\mathbf{X}; \mathbf{BC}) = \|\mathbf{X} - \mathbf{BC}\|_F^2$, where λ_1 is a constant. For the loss function, we choose a hinge squared loss function: $\ell(y, \tilde{y}) = (\max\{0, 1 - y\tilde{y}\})^2$, a common choice in Support Vector Machine (SVM) literature [3].

There are three blocks in the optimization problem in Eq. (1): \mathbf{w} , \mathbf{B} , and \mathbf{C} which is only jointly convex. In other words, if any two pairs of blocks, are fixed,

the problem is convex with respect to the remaining block. The optimization scheme starts from a random initialization of blocks, fixes two blocks, optimizes with respect to the remaining one, and repeats this process for each block. The whole process is repeated till convergence. Optimization with respect to \mathbf{C} and \mathbf{w} is not challenging but, due to the large-scale dimensionality of a medical image, optimization with respect to \mathbf{B} requires a specialized method (see [3] for details).

3 Extension to Multi-Modality

Unlike the uni-modality case, in which each voxel stores a scalar value, in the multi-modality case, each voxel of an image is associated with an array of values. In Section 2, we stored the training data into a matrix (\mathbf{X}); while in multi-modality case, we need to structure the data into a tensor (\mathbb{X}). In fact, in the general framework (Section 2), the matrix \mathbf{f} can be viewed as an order-2 tensor² in which the first index (rows) enumerates voxels and the second index (columns) enumerates subjects. We simply extend this matrix to an order-3 tensor in which the third index (faces) enumerates modalities. One can simply concatenate all image modalities of a subject into long columns of a matrix and simply apply [2] or a similar method. However, the advantage of extending a regularized matrix factorization to a tensor factorization framework is that various factorizations can be proposed each of which implies different hypotheses about the data because of the structure of a tensor. In this paper, we introduce two factorizations and explain their connotations (pictorially represented in Fig 1).

Our method can be viewed as *multi-view* learning [11]. In the multi-view setting, the goal is to implicitly learn about the target via the relationship between different views [11]. Depending on how to define targets, we can have different variations of the method. For example, if multiple modalities are different frequencies in spectroscopy imaging, different features extracted from diffusion tensor image (DTI), or time series in fMRI. One assumption could be that there is one hidden variable (here basis vectors: \mathbf{B}) that is shared across image modalities and class labels. Therefore, both class labels (\mathbf{y}) and data (\mathbb{X}) are the targets; we will refer to the method as **multi-View**(\mathbb{X}, \mathbf{y}).

Unlike **multi-View**(\mathbb{X}, \mathbf{y}), an alternative assumption could be that there is no hidden variable shared across modalities, hence every modality has its own basis vectors ($\mathbb{B}^{(v)}$), but projection on these basis vectors collaborate to predict class labels. For example, different modalities may measure quantities on non-overlapping regions of a brain (*e.g.* white matter and gray matter) each quantifying complementary features about the class labels. We refer to this variation as **multi-View**(\mathbf{y}). This assumption is still different than applying the uni-modality method separately because $\mathbb{B}^{(v)}$'s need to collaborate on the discriminative term.

² The order of a tensor is the number of indices necessary to refer unambiguously to an individual component of a tensor.

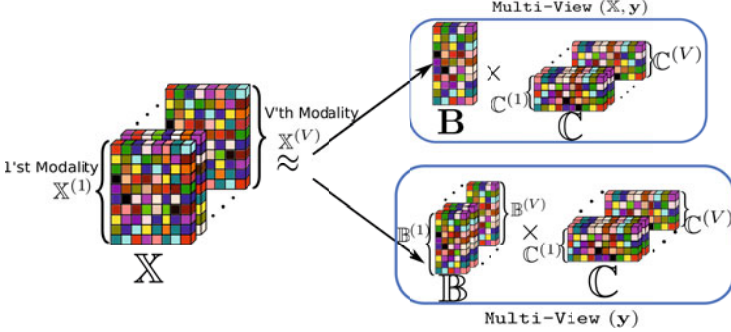


Fig. 1. The difference between the two proposed factorizations: **multi-View(\mathbf{y})** versus **multi-View(\mathbb{X}, \mathbf{y})**. There are V modalities stored in the data tensor (\mathbb{X}); for **multi-View(\mathbf{y})**, we need to have V sets of basis vectors ($\mathbb{B}^{(1)}, \dots, \mathbb{B}^{(V)}$) and corresponding coefficients ($\mathbb{C}^{(1)}, \dots, \mathbb{C}^{(V)}$), while for **multi-View(\mathbb{X}, \mathbf{y})**, there is one set of basis vectors (\mathbb{B}) shared across modalities.

The definitions of the generative term ($\mathcal{D}(\cdot; \cdot)$) and the classifier function ($f(\cdot)$) in Eq. (11) for tensor are changed accordingly to **multi-View(\mathbb{X}, \mathbf{y})** and **multi-View(\mathbf{y})** (depending on the assumptions on data):

$$\begin{array}{ll}
 \text{multi-View}(\mathbb{X}, \mathbf{y}): & \text{multi-View}(\mathbf{y}): \\
 \mathcal{D}(\mathbb{X}; \mathbb{B}, \mathbb{C}) = \sum_{v=1}^V \|\mathbb{X}^v - \mathbb{B}\mathbb{C}^v\|_F^2 & \mathcal{D}(\mathbb{X}; \mathbb{B}, \mathbb{C}) = \sum_{v=1}^V \|\mathbb{X}^v - \mathbb{B}^v\mathbb{C}^v\|_F^2 \\
 f(\mathbb{X}_n; \mathbf{W}, \mathbb{B}) = \sum_{v=1}^V \langle \mathbf{w}^v, \mathbb{B}^T \mathbb{X}_n^v \rangle & f(\mathbb{X}_n; \mathbf{W}, \mathbb{B}) = \sum_{v=1}^V \langle \mathbf{w}^v, (\mathbb{B}^v)^T \mathbb{X}_n^v \rangle
 \end{array}$$

where \mathbb{X} and \mathbb{C} are tensors of order-3 holding respectively all images and coefficients of the basis vectors. \mathbb{X}^v and \mathbb{C}^v are order-2 tensors (*i.e.* matrix) holding images and coefficients of v^{th} modality respectively. \mathbb{X}_n is a order-2 tensor holding all modalities of the n^{th} subject and \mathbb{X}_n^v is a order-1 tensor (*i.e.* vector) holding only v^{th} modality of the n^{th} subject. V is the number of modalities (views), $\langle \cdot, \cdot \rangle$ and $\|\cdot\|_F$ indicate inner product and Frobenius norm respectively. \mathbf{W} is a matrix holding parameters of the classifier function and \mathbf{w}^v is its v^{th} column corresponding to the v^{th} modality. Notice that in **multi-View(\mathbf{y})**, the generative term is separable for each modality but basis matrices (\mathbb{B}^v 's) are coupled together through the loss function ($\ell(\cdot, \cdot)$) in Eq. (11); therefore, it is different than applying the uni-modality algorithm (Section 2) separately and concatenating extracted features later for a classifier.

4 Experiments

We acquired a subset of images from a longitudinal brain imaging study for validation of our method. The objective of this choice was to investigate the longitudinal progression of changes in brain structure (MRI) and brain function ($[^{15}\text{O}]\text{-water PET-CBF}$) in relation to cognitive change in cognitively normal

older adults. We used slopes of CVLT³ score over the follow-up period as a measure of cognitive function to subdivide the entire cohort into two groups: top 20% (25 subjects) showing the highest cognitive stability (CN: cognitively normal), and bottom 20% (25 subjects) showing the most pronounced cognitive decline (CD: cognitively declining).

All T1-MR images used in this study were pre-processed according to [6] and registered to a template. Two volumetric tissue density maps [13] were formed for white matter (WM), gray matter (GM) regions. These maps quantify an expansion (or contraction) to the tissue applied by the transformation to warp the image to the template space.

Samples are divided into five folds and 4/5 of samples are used for training basis vectors (an example of which is shown in Fig.2); projections on these basis vectors are used as features and are fed to a SVM classifier.

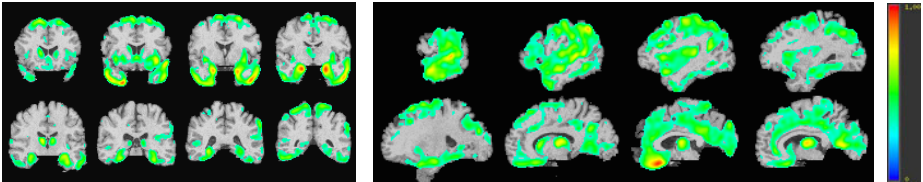


Fig. 2. Two examples of the basis vectors shown in different cuts. Left: $\text{Multi-View}(\mathbb{X}, \mathbf{y})$, Right: $\text{Multi-View}(\mathbf{y})$ ($\gamma^* = 100$; number of basis vectors is 60).

In uni-parametric dataset, the algorithm is relatively stable as long as λ 's are chosen within reasonable ranges (see [3]). We set the parameters to the most frequently chosen parameters used for a uni-modality case on a totally different dataset. Numbers reported in Table 1 are produced using such parameters. Nevertheless, we performed sensitivity analysis with respect to ratio of λ_1/λ_2 and number of basis vectors, K (see Fig.3). For notational brevity, we used γ^* for ratio of λ_1/λ_2 we used for Table 1. Different curves in Fig.3 denote different ratios of λ_1/λ_2 . As $\text{Multi-View}(\mathbf{y})$ is relatively stable with respect to K and different ratios, performance of $\text{Multi-View}(\mathbb{X}, \mathbf{y})$ improves as K increases. Although parameters that are more inclined toward the unsupervised setting (e.g. $\lambda_1/\lambda_2 = 10\gamma^*$) underperform settings that are excessively discriminative (e.g. $\lambda_1/\lambda_2 = 0.001\gamma^*$), are more stable. Weak regularization imposed on the excessively discriminative settings can explain this observation.

Table 1 reports the average classification rates on the left-out folds for different scenarios and methods. We used a publically available software, called COMPARE [6], for comparison. The COMPARE method has been applied to many problems and has been claimed to perform very well. Its variants, *i.e.* COMPARE and m-COMPARE, are similar to $\text{Multi-View}(\mathbf{y})$ and $\text{Multi-View}(\mathbb{X}, \mathbf{y})$ respectively. For comparison, we have included Single-View results for each scenario in which basis vectors are extracted independently and features are concatenated

³ California Verbal Learning Test [4].

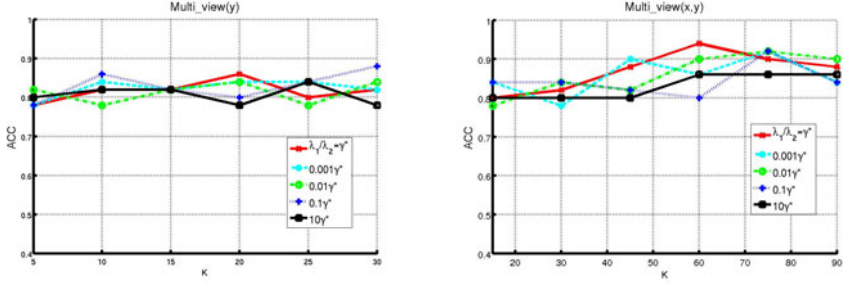


Fig. 3. Sensitivity Analysis: accuracy rates with respect to different number of basis vectors (K) for various ratios of λ_1/λ_2 . Left: Multi-View(\mathbf{y}). Right: Multi-View(\mathbb{X}, \mathbf{y})

and fed to the same procedure to find the best parameters for a classifier as the multi-view methods. Since results shown in the table are column-wise comparable, the highest values in the column are magnified with a bold font in each column. In general, Multi-View(\mathbb{X}, \mathbf{y}) or its counterpart m-COMPARE perform better. In all columns, at least one of the multi-view methods outperforms the single view equivalent and the best performance is achieved by Multi-View(\mathbb{X}, \mathbf{y}).

Table 1. Comparison of classification accuracy rates for different scenarios and different methods on “cognitively normal” (NC) versus “cognitively declining” (CD) subjects. Results are reported in the format: accuracy (sensitivity, specificity); with $\gamma^* = 100$; total number of basis vectors in each experiment is 60.

NC vs. CD				
	(WM,PET)	(WM,GM)	(GM,PET)	(GM, WM, PET)
Multi-View(\mathbb{X}, \mathbf{y})	0.82 (0.84,0.8)	0.76 (0.72,0.8)	0.84 (0.88,0.8)	0.94 (0.88,1.0)
Multi-View(\mathbf{y})	0.86 (0.84,0.88)	0.84 (0.8,0.88)	0.78 (0.8,0.76)	0.84 (0.84,0.84)
m-COMPARE	0.88 (0.8,0.96)	0.86 (0.88,0.84)	0.8 (0.8,0.8)	0.86 (0.84,0.88)
COMPARE	0.78 (0.68,0.88)	0.82 (0.76,0.88)	0.82 (0.84,0.8)	0.82 (0.76,0.88)
Single-View	0.84 (0.8,0.88)	0.84 (0.8,0.88)	0.82 (0.84,0.8)	0.8 (0.76,0.84)

5 Conclusion

We proposed a framework that exploits all modalities in a dataset simultaneously to reduce dimensionality in a discriminative yet interpretable way. Inspired by multi-view learning, two variants of constrained tensor factorization are suggested each of which implies different hypothesis about the data. We showed that the algorithm is relatively robust with respect to choice of parameters and achieves good classification results. Computational expense of the algorithm is moderate and as future work, we plan to apply it to case for which number of modalities is large (*e.g.* HARDI data or time series).

References

1. Ando, R.K., Zhang, T.: Two-view feature generation model for semi-supervised learning. In: Proceedings of the 24th International Conference on Machine Learning, ICML 2007, pp. 25–32. ACM, New York (2007)
2. Batmanghelich, N., Taskar, B., Davatzikos, C.: A general and unifying framework for feature construction, in image-based pattern classification. *Inf. Process. Med. Imaging* 21, 423–434 (2009)
3. Batmanghelich, N., Ye, D.H., Pohl, K., Taskar, B., Davatzikos, C.: Disease classification and prediction via semi-supervised dimensionality reduction. In: 2011 IEEE International Symposium on Biomedical Imaging (2011)
4. Delis, D., Kramer, J., Kaplan, E., Ober, B.: California Verbal Learning Test-Research Edition. The Psychological Corporation, New York (1987)
5. Diehl, J., Grimmer, T., Drzezga, A., Riemenschneider, M., Frstl, H., Kurz, A.: Cerebral metabolic patterns at early stages of frontotemporal dementia and semantic dementia. a pet study. *Neurobiol. Aging* 25(8), 1051–1056 (2004)
6. Fan, Y., Shen, D., Gur, R.C., Gur, R.E., Davatzikos, C.: Compare: classification of morphological patterns using adaptive regional elements. *IEEE Trans. Med. Imaging* 26(1), 93–105 (2007)
7. Foster, N.L., Heidebrink, J.L., Clark, C.M., Jagust, W.J., Arnold, S.E., Barbas, N.R., DeCarli, C.S., Turner, R.S., Koeppe, R.A., Higdon, R., Minoshima, S.: Fdg-pet improves accuracy in distinguishing frontotemporal dementia and alzheimer's disease. *Brain* 130(Pt 10), 2616–2635 (2007)
8. Fox, N.C., Schott, J.M.: Imaging cerebral atrophy: normal ageing to alzheimer's disease. *Lancet* 363(9406), 392–394 (2004)
9. Golland, P., Grimson, W.E.L., Shenton, M.E., Kikinis, R.: Deformation analysis for shape based classification. In: Insana, M.F., Leahy, R.M. (eds.) IPMI 2001. LNCS, vol. 2082, pp. 517–530. Springer, Heidelberg (2001)
10. Hinrichs, C., Singh, V., Xu, G., Johnson, S.: Mkl for robust multi-modality ad classification. *Med. Image Comput. Comput. Assist. Interv.* 12(Pt 2), 786–794 (2009)
11. Kakade, S., Foster, D.: Multi-view regression via canonical correlation analysis, pp. 82–96 (2007)
12. Landau, S.M., Harvey, D., Madison, C.M., Reiman, E.M., Foster, N.L., Aisen, P.S., Petersen, R.C., Shaw, L.M., Trojanowski, J.Q., Jack, C.R., Weiner, M.W., Jagust, W.J., Initiative, A.D.N.: Comparing predictors of conversion and decline in mild cognitive impairment. *Neurology* 75(3), 230–238 (2010)
13. Shen, D., Davatzikos, C.: Very high-resolution morphometry using mass-preserving deformations and hamper elastic registration. *Neuroimage* 18(1), 28–41 (2003)
14. Zhang, D., Wang, Y., Zhou, L., Yuan, H., Shen, D.: The ADNI: Multimodal classification of alzheimer's disease and mild cognitive impairment. *Neuroimage* (January 2011)

Detection, Grading and Classification of Coronary Stenoses in Computed Tomography Angiography

B. Michael Kelm¹, Sushil Mittal^{2,3,*}, Yefeng Zheng², Alexey Tsymbal¹,
Dominik Bernhardt⁴, Fernando Vega-Higuera⁴, S. Kevin Zhou²,
Peter Meer³, and Dorin Comaniciu²

¹ Image Analytics and Informatics, Corporate Technology, Siemens AG, Erlangen, Germany

² Image Analytics and Informatics, Siemens Corporate Research, Princeton, NJ, USA

³ Electrical and Computer Engineering, Rutgers University, NJ, USA

⁴ Computed Tomography, Healthcare Sector, Siemens AG, Forchheim, Germany

Abstract. Recently conducted clinical studies prove the utility of Coronary Computed Tomography Angiography (CCTA) as a viable alternative to invasive angiography for the detection of Coronary Artery Disease (CAD). This has led to the development of several algorithms for automatic detection and grading of coronary stenoses. However, most of these methods focus on detecting calcified plaques only. A few methods that can also detect and grade non-calcified plaques require substantial user involvement. In this paper, we propose a fast and fully automatic system that is capable of detecting, grading and classifying coronary stenoses in CCTA caused by *all* types of plaques. We propose a four-step approach including a learning-based centerline verification step and a lumen cross-section estimation step using random regression forests. We show state-of-the-art performance of our method in experiments conducted on a set of 229 CCTA volumes. With an average processing time of 1.8 seconds per case after centerline extraction, our method is significantly faster than competing approaches.

1 Introduction

According to the American Heart Association, Coronary Artery Disease (CAD) is a leading cause of death in the western world. Every year, about six million patients in the United States emergency departments are examined for acute chest pain [5]. The current diagnostic standard is conventional invasive angiography which involves a very high amount of risk and cost. New generations of cardiac Computed Tomography (CT) scanners enable the acquisition of Coronary CT Angiography (CCTA) images with unprecedented quality [2]. In the review article [1], Achenbach has summarized the results of many clinical studies, comparing contrast-enhanced CCTA with conventional invasive angiography. These results prove CCTA a viable alternative with very high negative predictive value. However, reading CCTA images requires substantial experience and only well-trained physicians are able to interpret CCTA reliably [12]. An automated system that can rule out clinically-relevant stenoses (grade > 50%) in the coronary arteries could be used as a second reader in the absence of an expert physician in the emergency department.

* The author has been with Siemens Corporate Research for this work.

Detection and grading of coronary stenoses in CCTA is very challenging due to varying image quality. In an endeavor to reduce the radiation dose during CT scans, often images with relatively low signal-to-noise ratio are acquired [2]. Motion artifacts are frequently encountered despite the routine use of beta blockade to reduce the heart rate. Gated or modulated acquisition protocols may result in stair-case reconstruction artifacts which further complicate the analysis. Finally, coronary arteries follow long and treacherous paths, extending over only a few voxels in diameter in the distal parts. Even experts sometimes struggle to give a correct diagnosis due to these challenges [12].

In this work, we present an algorithm for detecting, grading and classifying severe (i.e. clinically relevant) stenoses along automatically extracted centerlines of the coronaries. The contribution of our work is three-fold. First, we introduce a novel centerline verification step (Section 3.1). Second, a novel regression approach replaces the lumen segmentation that is needed for grade estimation (Section 3.2). Finally, in contrast to previous work, we propose a complete end-to-end and fully automated system that works on *all* types of plaques.

2 Related Work

Vessel Tracing. Manual tracing of coronary centerlines in 3D cardiac CT volumes is a highly tedious task. Many algorithms for automatic tracing of centerlines have been proposed, the most important of which are reviewed in [8]. Owing to the importance of the problem, the MICCAI association also organized a competition for automatic and semi-automatic coronary artery tracking [10]. In our work, the centerlines were traced using the state-of-the-art method of [6]. The method uses multi-scale *medialness* filters in a graph-based algorithm to extract centerlines by computing minimum-cost paths.

Lumen Segmentation. Estimating the cross-sectional area (or the radius) of the vessel lumen along the centerline is a key feature to detecting and grading coronary stenoses. But methods relying on exact segmentation of the lumen are slow owing to high computational complexity of the segmentation algorithms. A comprehensive review of all major lumen segmentation algorithms is provided in [8]. Given a centerline, we propose a novel, automatic regression-based method to directly estimate the vessel radius, which is significantly faster than a segmentation-based approach.

Stenosis Detection. In the past, a variety of algorithms have been proposed for detection of coronary plaques in CCTA. However, most of this work focuses on the detection of calcified plaques only, e.g., see [14,11]. Fewer methods have been proposed for fully automatic detection of non-calcified plaques, which are usually harder to detect and grade with high confidence. For example, the methods proposed in [13] and [15] need substantial user input in order to localize and grade coronary stenoses. A learning based method was proposed in [14] which could detect both calcified and non-calcified plaques. However, this approach does not consider stenosis grade and thus also reports non-severe lesions which are clinically irrelevant. To the best of our knowledge, the only system that currently analyzes CCTA for coronary artery disease in a fully automated way by reporting location, type and severity of coronary lesions is Rcadia's COR

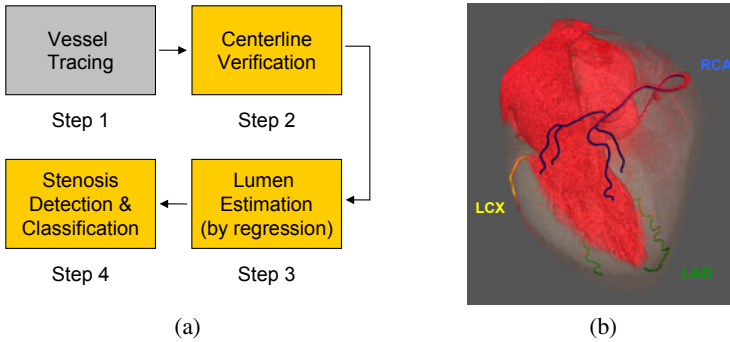


Fig. 1. (a) Overview of the four-step approach for stenosis detection, grading and classification. (b) Example vessel tree showing the three major coronary arteries: left anterior descending (LAD), left circumflex (LCX) and right coronary artery (RCA).

Analyzer. Being a commercial product, however, not much detail is known about the approach. Recent studies evaluating the system were published in [3] and [7].

3 Methods

Given a CCTA image volume, the proposed system can automatically detect and classify coronary stenoses using the general four-step approach sketched in Fig. 1a. In the *first* step, centerlines of the three major coronary arteries, i.e. left anterior descending (LAD), left circumflex (LCX) and right coronary artery (RCA) (Fig. 1b), along with their branches, are automatically extracted using the method of [6]. The left main (LM) coronary artery is processed as the common part of LAD and LCX. Using a learning-based classification approach, the *second* step verifies the accuracy of the extracted centerlines and removes parts of the vessel tree belonging to non-coronary regions. This ensures that the subsequent stenosis detection is only performed along the actual vessel, thus producing more stable results. The *third* step employs a learning-based regression approach to locally estimate the cross-sectional area of the vessel lumen (which may be narrowed by coronary plaques), along the extracted centerlines. In the *fourth* step, candidate stenoses are extracted for each individual segment of the vessel tree as max-min-max triples of a baseline-corrected and smoothed radius curve along with a grade estimate. Each candidate stenosis is either discarded (e.g. grade < 50%) or accepted and classified as calcified, non-calcified or mixed stenosis. This decision is based on image features as well as features of the candidate stenosis such as length, cross-sectional area/radius and distance to the distal end of the vessel.

3.1 Centerline Verification

To avoid the risk of missing a potentially diseased coronary vessel, any automatic centerline extraction algorithm must be highly sensitive, especially in low-contrast vessel regions, which is true for the employed algorithm [6], for example. Thus, the method should be able to trace centerlines of the vessels occluded by non-calcified plaques, that

often exhibit the same contrast as the vessel wall. Inevitably, such algorithms are prone to various types of errors in tracing and centerlines may get wrongly traced into other arteries or into non-coronary regions like veins, heart chambers, etc. Figure 2 (left) shows Curved Planar Reformation (CPR) views of some more noticeable examples for erroneous tracing results of the employed algorithm. Table 1 summarizes the errors in tracings for 229 volumes with 1472 traced centerlines in the distal vessel regions.

Table 1. Summary of errors in tracing coronary centerlines in the distal vessel regions. A total of 229 volumes were processed and 1472 centerlines were traced.

Error in tracing	$\geq 5\text{mm}$	$\geq 10\text{mm}$	$\geq 15\text{mm}$
# vessels affected	259 (17.6%)	226 (15.4%)	210 (14.3%)
# volumes affected	131 (57.2%)	116 (50.66%)	107 (46.7%)

Clearly, erroneously traced centerlines are susceptible to many false alarms while detecting lesions. Therefore, we propose to use a learning based algorithm for automatic detection of non-coronary regions along the extracted centerlines. Similar to [11], a cylindrical sampling pattern for feature extraction, with its axis aligned to the coronary centerline, is employed. We then extract altogether 171 rotation invariant features along the entire length of the cylinder at varying radii. These features are used to train a random forests (RF) classifier [4] (100 trees, 14 randomly selected features, stratified sampling of 8000 examples with replacement). Given an unseen volume, the RF classifier outputs a probability that a given point on the centerline belongs to a non-vessel region. After removing outliers using a median filter of width 11, points with scores higher than a fixed threshold are excluded from further analysis. In our experiments this threshold was determined as 0.7 to yield a specificity of at least 98%, so that at most 2% of the points are erroneously discarded (see Fig. 2 right).

3.2 Lumen Estimation

Instead of segmenting the lumen and computing the lumen cross-sectional area along the vessel centerlines we propose using a non-linear regression approach to directly estimate the cross-sectional area from local image features.

Alternatively for estimating the cross-sectional area we trained for and estimated the radius r of an equivalent circle with the same area. Thus, a function for the radius $r(\mathbf{x}|p)$ is estimated that depends on the feature vector \mathbf{x} and a set of parameters p that are learned from a training set $T = \{(\mathbf{x}_i, r_i)\}_{i=1}^N$. The training set is constructed from semi-manual lumen segmentations of coronary arteries by computing cross-sections and the corresponding radii r_i at altogether N centerline points. At the same points, rotation-invariant features \mathbf{x}_i are extracted according to [11]. A regression function is learned by minimizing the squared loss $L(p) = \sum_{i=1}^N (r(\mathbf{x}_i|p) - r_i)^2$ on the training set T with respect to the parameters p . For this purpose an ensemble of randomized regression trees, i.e. a random regression forest was employed [4]. Formally, each regression tree models the dependent variable (the lumen radius) as a piecewise constant function, which results in the following model for the ensemble:

$$r(\mathbf{x}|p) = \frac{1}{T} \sum_{t=1}^T \sum_{m=1}^{M^t} r_m^t I(\mathbf{x} \in \mathbf{R}_m^t) \quad (1)$$

where $I(\cdot)$ denotes the set indicator function. Thus, for each tree t , the feature domain is partitioned into M^t regions $\mathbf{R}_1^t, \mathbf{R}_2^t, \dots, \mathbf{R}_{M^t}^t$, and in each region the radius is modeled by a constant r_m^t [4]. During training of each tree, the sample is successively partitioned by thresholding one of the feature values. The optimal split (feature and threshold) is greedily determined as the one which reduces the squared loss function most. In the leaves (a region $\mathbf{R}_{M^t}^t$), the radius r_m^t is estimated as the average radius of all remaining examples. While it is possible to stop splitting when the variance within a node and/or the number of examples drop below certain values, we use fully grown trees in our experiments, i.e. they were grown until only one example remained in every leaf. For obtaining good generalization performance, it is essential that every tree slightly differs. This is achieved by training each tree on a resampled training sample (at most 4000 examples drawn with replacement) and by randomly selecting a fraction of the features (one third, i.e. 57 of 171) to be considered for each split.

Given the feature vector of an unseen example, each regression tree is evaluated by following its splits from the root to a leaf with an associated radius estimate r_m^t . The random forest averages the estimates of its trees (Eqn. (1)). We used $T = 50$ regression trees in our experiments.

3.3 Stenosis Detection and Classification

Candidate stenosis regions are identified and graded using the estimated lumen radii. The vessel tree is decomposed into disjoint segments so that every segment either starts at an ostium or a vessel bifurcation. Each segment is then analyzed separately.

First, a baseline curve is computed using binomial filtering (Fig. 3 green). It is subtracted from the original radius curve to obtain a de-trended residual curve which is again slightly smoothed (Fig. 3 red). The positions of the local optima are extracted; clearly, local minima and maxima alternate. Every triple (max-min-max) is then regarded as a stenosis candidate for which a grade is estimated by

$$g = 1 - \left(\frac{2r_{min}}{r_{left} + r_{right}} \right)^2, \quad (2)$$

where r_{min} is the minimum radius within, r_{left} the radius at the left (towards the ostium) and r_{right} the radius at the right end of the stenosis candidate (Fig. 3 magenta). At the ostia and bifurcations, the grade is estimated with the alternative formula $g = 1 - (r_{min}/r_{right})^2$ to account for the non-pathologic radius broadening there.

Then it is decided for each stenosis candidate whether it should be discarded or not. In addition to low grade candidates ($< 50\%$), also short (< 0.9 mm) and narrow ($r_{right} < 1.0$ mm) ones are discarded. Candidates close to the distal end of the vessel (< 7.5 mm) are also discarded since the data quality usually gets too low there.

Finally, using probability scores obtained from two classifiers for the detection of calcified and non-calcified plaques similar to [11] (Fig. 3 cyan/orange), each accepted

stenosis candidate is classified into one of three types, “calcified”, “non-calcified” and “mixed” (calcified as well as non-calcified parts).

4 Experimental Results

Training of the system was performed using a total of 229 CCTA volumes that were acquired on several cardiac CT scanners with varying protocols and reconstruction algorithms. The slice distance for these scans varied between 0.3-0.5mm with x - y pixel spacing being between 0.3-0.4mm. Each scan typically consisted of 200-300 slices. For training and evaluation, the data was manually annotated. For automatic vessel tracing, errors were annotated. Coronary plaques were labeled with their type (calcified, non-calcified, mixed) as well as a rough grade (mild, moderate, severe, occluded).

Figure 2 shows cross-validation results for the centerline verification step. Invalid centerline points are reliably recognized for several types of tracing errors (Fig. 2 *Left*). The receiver-operating-characteristic (ROC) curve (Fig. 2 *Right*) shows that high specificities as well as sensitivities are attainable. The inferior performance on the RCA results from the difficulty to distinguish the (invalid) coronary sinus vein from a (valid) artery. For the centerline verification step a high specificity (> 0.98) is desirable in order to ensure that no valid parts of the vessel tree are discarded.

Numeric results for the overall system are provided in Table 2. The “by-lesion” sensitivity quantifies how many of the severe non-calcified stenoses are detected while the “by-vessel” sensitivity quantifies how many of the vessels with severe non-calcified stenoses are identified. Only the “by-vessel” measure allows to compute the specificity and negative predictive value (NPV). For an application as a second reader, sensitivity and NPV are of utmost importance and preferred over specificity, i.e. false positives are acceptable while false negatives are not. With an overall sensitivity of 97.62%, a NPV of 99.77% and a specificity of 67.14% the proposed system performs competitive with

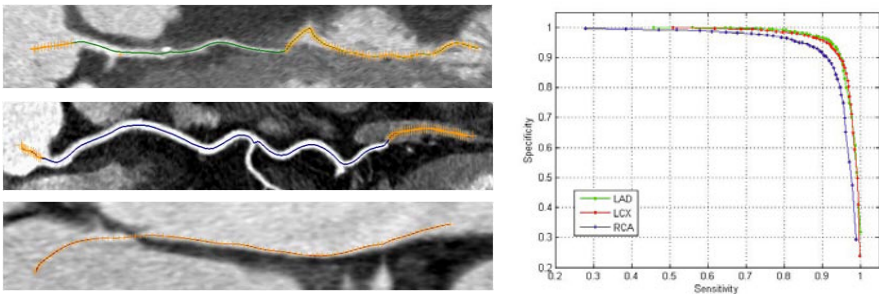


Fig. 2. *Left:* CPR views of wrong centerline tracings obtained using the algorithm of [6]. The first two are partially incorrect and traced into a heart chamber and a vein respectively. The third is entirely wrong and traced into a heart chamber. Points detected outside the coronary are marked with orange ‘+’ sign. *Right:* ROC curves (sensitivity vs. specificity) obtained with 10-fold cross validation using 229 volumes on per vessel point basis.

Table 2. Detection performance on non-calcified plaques (10-fold cross-validation). The system is tuned towards high sensitivity and NPV to cope with its application as a second reader.

		LAD	LCX	RCA	Overall
by-lesion	sensitivity	100.0%	90.0%	95.24%	94.55%
	false positives per volume	0.81	1.03	1.13	2.97
by-vessel	sensitivity	100.00%	93.75%	100.00%	97.62%
	specificity	75.23%	63.16%	62.86%	67.14%
	negative predictive value	100.00%	99.17%	100.00%	99.77%

Rcadia’s system [37]. But, while the Rcadia system requires about 10min of processing time per case, on average, our system only needs 1.8 s (up to 3.9 s) after centerline tracking which completes within a minute.

This advantage can mainly be attributed to excellent performance of the lumen regression step, which provides an accurate estimate of the lumen cross-sectional area much faster than a segmentation method. To this end, we also compared our results to a segmentation approach, a learning-based version of the graph cuts approach presented in [9]. While comparable performance was achieved, the overall processing time of 21 s (up to 42 s) after centerline extraction was clearly higher with segmentation approach.

A test of the lumen regression (without retraining) on the eight (training) data sets provided by [10] yielded a bias of 0.18 mm (median 0.19 mm) and a standard deviation of 0.27 mm (median absolute deviation 0.14 mm). Apart from the bias, which can be attributed to systematic annotation differences, these results agree with the 10-fold cross-validation results on our data which yielded a bias of 0.01 mm (median 0.002 mm) and a standard deviation of 0.28 mm (median absolute deviation 0.08 mm).

Fig. 3 shows two examples of patients that have neither been used in training nor the development of the proposed system. In both cases, the system can locate and classify all severe stenotic lesions correctly. Note that the right example shows a second, more distal stenosis for which a grade of about 40% is estimated (cf. graph below the image). It is thus deemed non-severe and therefore not reported.

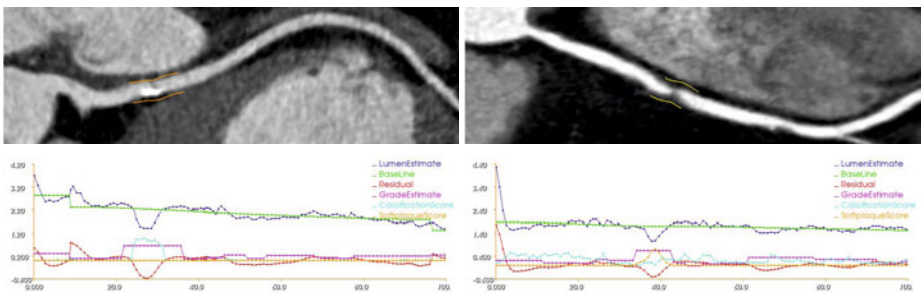


Fig. 3. Two examples of patients from unseen data. The presented system detects severe stenoses caused by both calcified plaques (left) and non-calcified plaques (right). The graph at the bottom shows the lumen estimate (blue), baseline (green), residual (red), grade estimate (magenta), calcification score (cyan) and the score for non-calcified (soft) plaques (orange) along the vessel centerline. While the system is highly sensitive it only exhibits a moderate number of false alarms. For the two examples shown above, other severe stenoses were neither present nor detected.

5 Conclusion

An automatic method for the detection and classification of stenotic lesions in coronary computed tomography angiography is proposed. The centerline verification step helps the system cope with tracing errors and vessels with low data quality. Lumen cross-sectional area is accurately estimated using a regression approach which is considerably faster than a full segmentation method. A competitive performance is achieved with significantly reduced computational burden as compared to state-of-the-art methods.

References

1. Achenbach, S.: Cardiac CT: State of the art for the detection of coronary arterial stenosis. *J. Cardiovasc. Comput. Tomogr.* 1(1), 3–20 (2007)
2. Achenbach, S., Anders, K., Kalender, W.: Dual-source cardiac computed tomography: Image quality and dose considerations. *Eur. Radiol.* 18, 1188–1198 (2008)
3. Anders, K., Petit, I., Achenbach, S., Pflederer, T.: Diagnostic utility of automated stenosis detection in dual source CT coronary angiography as a stand alone or add-on tool. In: *Proc. SCCT (2010)*
4. Breiman, L.: Random forests. *Mach. Learn.* 45(1), 5–32 (2001)
5. Gallagher, M.J., Raff, G.L.: Use of multislice CT for the evaluation of emergency room patients with chest pain: the so-called “triple rule-out”. *Cath. and Cardiovasc. Interv.* 71(1), 92–99 (2008)
6. Gülsün, M.A., Tek, H.: Robust vessel tree modeling. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) *MICCAI 2008, Part I. LNCS*, vol. 5241, pp. 602–611. Springer, Heidelberg (2008)
7. Halpern, E.J., Halpern, D.J.: Diagnosis of coronary stenosis with CT angiography comparison of automated computer diagnosis with expert readings. *Acad. Radiol.* 18, 324–333 (2011)
8. Lesage, D., Angelini, E.D., Bloch, I., Funka-Lea, G.: A review of 3D vessel lumen segmentation techniques: Models, features and extraction schemes. *Med. Image Anal.* 13(6), 819–845 (2009)
9. Li, K., Wu, X., Chen, D.Z., Sonka, M.: Optimal surface segmentation in volumetric images—a graph-theoretic approach. *IEEE Trans. Pattern Anal. Mach. Intell.* 28(1), 119–134 (2006)
10. Metz, C., Schaap, M., van Walsum, T., van der Giessen, A., Weustink, A., Mollet, N., Krestin, G., Niessen, W.: 3D segmentation in the clinic: A grand challenge II – coronary artery tracking. In: *MICCAI Workshop on 3D Segmentation in the Clinic: A Grand Challenge (2008)*
11. Mittal, S., Zheng, Y., Georgescu, B., Vega-Higuera, F., Zhou, S., Meer, P., Comaniciu, D.: Fast automatic detection of calcified coronary lesions in 3D cardiac CT images. In: *Proc. MICCAI Workshop on Machine Learning in Medical Imaging (2010)*
12. Pugliese, F., Hunink, M.G.M., Gruszczynska, K., Alberghina, F., Malag, R., van Pelt, N., Mollet, N.R., Cademartiri, F., Weustink, A.C., Meijboom, W.B., Witteman, C.L.M., de Feyter, P.J., Krestin, G.P.: Learning Curve for Coronary CT Angiography: What Constitutes Sufficient Training? *Radiol.* 251(2), 359–368 (2009)
13. Rinck, D., Krüger, S., Reimann, A., Scheuering, M.: Shape-based segmentation and visualization techniques for evaluation of atherosclerotic plaques in coronary artery disease. In: *Proc. SPIE Int. Soc. Opt. Eng.*, vol. 6141, pp. 61410G–9 (2006)
14. Teßmann, M., Vega-Higuera, F., Fritz, D.: Learning-based detection of stenotic lesions in coronary CT data. In: *Proc. of Vision, Modeling, and Visualization*, pp. 189–198 (2008)
15. Wesarg, S., Khan, M.F., Firle, E.: Localizing calcifications in cardiac CT data sets using a new vessel segmentation approach. *J. of Dig. Imag.* 19(3), 249–257 (2006)

Aggregated Distance Metric Learning (ADM) for Image Classification in Presence of Limited Training Data

Gaoyu Xiao and Anant Madabhushi*

Department of Biomedical Engineering, Rutgers University, USA
{gyxiao, anantm}@rci.rutgers.edu

Abstract. The focus of image classification through supervised distance metric learning is to find an appropriate measure of similarity between images. Although this approach is effective in the presence of large amounts of training data, classification accuracy will deteriorate when the number of training samples is small, which, unfortunately, is often the situation in several medical applications. We present a novel image classification method called *aggregated distance metric* (ADM) learning for situations where the training image data are limited. Our approach is novel in that it combines the merits of boosted distance metric learning (BDM, a recently published learning scheme) and bagging theory. This approach involves selecting several sub-sets of the original training data to form a number of new training sets and then performing BDM on each of these training sub-sets. The distance metrics learned from each of the training sets are then combined for image classification. We present a theoretical proof of the superiority of classification by ADM over BDM. Using both clinical (X-ray) and non-clinical (toy car) images in our experiments (with altogether 10 sets of different parameters) and image classification accuracy as the measure, our method is shown to be more accurate than BDM and the traditional bagging strategy.

1 Introduction

Image classification is important in many medical applications, for instance, to distinguish images representing different pathologies in the context of content-based image retrieval (CBIR). Another example is to identify common anatomical landmarks for the purpose of image data fusion and image registration. In general, medical image classification can be explored in two fronts: (1) extracting a representative set of features and (2) finding an appropriate similarity measure between images. The latter, named distance metric learning, is not as extensively explored as the former, it therefore has great potential to further improve the image classification accuracy [1]. Most distance metric learning (DML) methods can be classified as either unsupervised or supervised. The supervised

* Thanks to funding agencies: National Cancer Institute (R01CA136535-01, R01CA140772 01, R03CA143991-01), and The Cancer Institute of New Jersey.

approaches, which require training images with user-defined labels or pair-wise constraints, are more frequently found in medical applications [2].

Boosting framework [3], which uses a set of weak learners to create a strong learner, has been adopted in DML. Recently, a boosted distance metric learning method (BDM) using pair-wise constraints was proposed [1]. This method, when combined with the nearest neighbor search, has been proven to be efficient and accurate in classifying medical images of multiple classes when size of the training data is large enough [1].

However, well annotated training data are almost always difficult to obtain in medical imaging problems. In particular, in supervised multi-class image classification, there may be only limited number of training images for each image class (e.g. a brain atlas can contain a large number of anatomical landmarks derived from a much smaller number of subjects), hence the sample size for each class may not be statistically reasonable [4] [5]. One way to improve the performance of image classification in the presence of limited training images may be via bootstrap aggregating (bagging) [6], wherein bootstrapped training sets are constructed and classification is then performed on each set. The final classification is achieved through a plurality vote. However, when the performance of each individual classifier is constrained by the size of the training data, further improvement will be desirable.

In this work, we present a novel method called *aggregated distance metric learning* (ADM) to classify a test image, specifically in the context of (a) large number of classes, and (b) limited training images for each class. Inspired by the idea of BDM [1] and bagging [6], instead of using the whole training image set Ω of k classes C_i , ($i \in \{1, \dots, k\}$) to derive a single distance metric d , our method first selects M sub-sets Ω_m ($m \in \{1, \dots, M\}$) from Ω , then performs BDM on each Ω_m to obtain a unique distance metric d_m . In order to determine which class C' a test image t belongs to, the distance between t and each C_i is computed using every distance metric $d_m(t, C_i)$. Next, all d_m are aggregated $\sum_{m=1}^M d_m(x, C_i)$ and the class C' with the smallest aggregated distance is identified as the class label. It can be seen that ADM is a meta-algorithm based on BDM. Moreover, it differs from bagging in that bagging conducts a plurality vote on all classification results of each predictor, while ADM computes the aggregated distances from all d_m to get the classification result.

To our knowledge, there is no previous work trying to combine the merits of DML and bagging in the context of image classification. In this work, we present a rigorous theoretical analysis to show why ADM yields better classification compared to BDM. In addition, our method is more accurate than the traditional bagging approach, as the continuous aggregated distance value is more robust against the errors caused by small training set sizes. We demonstrate the superiority of ADM over other state of the art methods in experiments involving both clinical (X-ray) and non-clinical (toy car) image data. We also demonstrate the potential applicability of ADM in the context of CBIR applications.

2 Previous Related Work

2.1 Boosted Distance Metric Learning

Yang et al proposed BDM for image retrieval and classification [1], where the distance function between data points x_1 and x_2 was defined as:

$$d(x_1, x_2) = \sum_{i=1}^T \alpha_i (f_i(x_1) - f_i(x_2))^2, \quad (1)$$

where $f_i(x)$ is a binary classification function, and α_i are the combination weights, and T is the number of iterations. For the specific task of image classification, x_1 and x_2 may present images or extracted features. With the pair-wise constraints, the appropriate classification function $f_i(x)$ and the combination weights α_i can be learned by using the bound optimization theory [1].

During image classification, all training images are used to learn distance metric d , which is then used to compute the distance between a test image t and each training image. Based on these distance values, the class that is closest to t is considered as the class that the test image should belong to.

2.2 Bootstrap Aggregating (Bagging)

In [6] Breiman showed that classification accuracy can be improved by bootstrap sampling of the original data set to generate multiple versions of a classifier, and then using the results from the individual classifiers to obtain a consensus prediction. A theoretical proof was provided in [6] to explain why the bagging classifier, on average, performs better compared to the individual classifiers.

3 Classification by Aggregated Distance Metric Learning (ADM)

3.1 Theoretical Intuition

Inspired by the work of BDM [1] and bagging [6], ADM integrates the best of both approaches. It first constructs a number of training image sub-sets, then obtains the aggregated distance from these sub-sets for image classification.

3.2 Constructing Training Image Sub-sets

We construct M sub-sets of training images from the original training data set Ω . Suppose there are altogether k different classes of training images C_1, \dots, C_k , and each class C_i , ($i \in \{1, \dots, k\}$) consists of n training images $C_i = \{(x_{ij}, y_i)\}$, ($i \in \{1, \dots, k\}; j \in \{1, \dots, n\}$), where x_{ij} is a training image and y_i is the corresponding class label. Since in this work we are interested in the scenario where the number of training images for each class, n , is small, we assume $k > n$. At each iteration, we randomly choose n classes of training images to form a training image sub-set Ω_m so that we have a total of $M = \binom{k}{n}$ training image sub-sets Ω_m , ($m \in \{1, \dots, M\}$), where Ω_m consists of all the training images in the n classes that have been chosen.

3.3 ADM for Image Classification

BDM is performed on each sub-set Ω_m to obtain corresponding distance metric $d_m(x_1, x_2)$ between two images x_1 and x_2 , ($m \in \{1, \dots, \binom{k}{n}\}$). Given a test image t , the aggregated distance between t and a class $C_i = \{(x_{ij}, y_i)\}$, ($i \in \{1, \dots, k\}$; $j \in \{1, \dots, n\}$) using d_m is defined as:

$$d_m(t, C_i) = \frac{1}{n} \sum_{j=1}^n d_m(t, x_{ij}). \quad (2)$$

Finally, by summing up all d_m , ($m \in \{1, \dots, M\}$), $M = \binom{k}{n}$, we get the *aggregated distance* between image t and class C_i as

$$D(t, C_i) = \sum_{m=1}^M d_m(t, C_i), \quad (3)$$

so that the class C' with the smallest aggregated distance is chosen as the class which t should belong to. Mathematically, this is formulated as:

$$C' = \arg \min_i [D(t, C_i)]. \quad (4)$$

Next, we will prove that classification by ADM is better than BDM.

3.4 Proof of Superiority of Classification by ADM over BDM

We give a mathematical proof to show that classification by ADM is more accurate compared to BDM.

Preliminaries. We denote the probability that a test image t is classified into class C using the m -th distance learning process as $\psi_m(t, C)$, $m \in \{1, \dots, M\}$ ($M = \binom{k}{n}$) if the training image sub-set are constructed according to Section 3.2). We also denote the probability that t is classified into class C using all the distance learning processes (ensemble classifier) as $\Psi(t, C)$.

Theorem. Given that the correct classification of t is class C_r , then $\Psi(t, C_r) \geq \psi_m(t, C_r)$.

Proof. Given that a wrong classification of t is class C_w , then statistically, it is reasonable to assume

$$E(d_m(t, C_w)) > E(d_m(t, C_r)), \quad (5)$$

where $E()$ is the expectation operation. We introduce an auxiliary variable

$$g_m^t = d_m(t, C_w) - d_m(t, C_r). \quad (6)$$

Let $E(g_m^t) = \mu$ and $var(g_m^t) = \sigma^2$. According to Eq. 5, we have $\mu > 0$.

The probability that t is closer to C_r when measured by the aggregated distance is

$$P(D(t, C_w) > D(t, C_r)) = P\left(\sum_{m=1}^M g_m^t > 0\right). \quad (7)$$

According to central limit theorem, the distribution of $\sum_{m=1}^M g_m^t$ approaches a normal distribution $N(M\mu, \sigma^2)$ as M becomes larger (since $E(g_m^t) = \mu$ and $\text{var}(g_m^t) = \sigma^2$), so that

$$E\left(\sum_{m=1}^M g_m^t\right) = M\mu; E(g_m^t) = \mu. \quad (8)$$

Since given two distributions with the same variance, the one with a larger expectation has a larger probability of being positive, we have

$$P\left(\sum_{m=1}^M g_m^t > 0\right) \geq P(g_m^t > 0). \quad (9)$$

Expanding Eq. 9 using Eq. 6, we have

$$P\left(\sum_{m=1}^M d_m(t, C_i) > \sum_{m=1}^M d_m(t, C_r)\right) \geq P(d_m(t, C_i) > d_m(t, C_r)), \text{ where } i \in \{1, \dots, r-1, r+1, \dots, k\} \quad (10)$$

Since probability values are non-negative, from Eq. 10 we have

$$\prod_{i \neq r, i=1}^k (P\left(\sum_{m=1}^M d_m(t, C_i) > \sum_{m=1}^M d_m(t, C_r)\right)) \geq \prod_{i \neq r, i=1}^k (P(d_m(t, C_i) > d_m(t, C_r))). \quad (11)$$

According to the definitions of $\psi_m(t, C_r)$ and $\Psi(t, C_r)$, we have

$$\psi_m(t, C_r) = \prod_{i \neq r, i=1}^k P(d_m(t, C_i) - d_m(t, C_r) > 0), \quad (12)$$

and

$$\Psi(t, C_r) = \prod_{i \neq r, i=1}^k P\left(\sum_{m=1}^M d_m(t, C_i) - \sum_{m=1}^M d_m(t, C_r) > 0\right), \quad (13)$$

by combining Eqs 11, 12 and 13, we have

$$\Psi(t, C_r) \geq \psi_m(t, C_r). \square \quad (14)$$

4 Experimental Results and Discussion

We compared image classification accuracy of ADM, BDM and bagging. For the bagging approach, BDM classification was done for each sub-set before a plurality vote. Both clinical and non-clinical images were used. We directly used image pixel intensities in the metric learning, so that the results obtained in this way would not depend on any specific image feature. We also performed preliminary experiments on the potential application of ADM in image retrieval. In order to do so, following the notation in Eq 3, we define the aggregated distance between t and x_{ij} as $D(t, x_{ij}) = \sum_{m=1}^M d_m(t, x_{ij})$ and employ this measure for image retrieval.

Table 1. Comparison of classification accuracy of clinical images using different methods (ADM, BDM and bagging). The highest accuracy values are shown bolded.

Number of classes (k)	Number of training image per class (n)	Number of test images	Classification accuracy		
			ADM	BDM	Bagging
6	3	68	0.50	0.31	0.38
10	3	106	0.39	0.12	0.22
8	4	80	0.43	0.27	0.38
14	4	128	0.39	0.11	0.25
10	5	86	0.43	0.30	0.36
10	6	76	0.69	0.66	0.68

4.1 Experiments Using Clinical Images

The clinical images used are the X-ray images from ImageCLEFmed2009 data set¹, which consists of 12677 images that have already been categorized into 193 classes by experts. We randomly selected images of k classes, and for each class, n images were randomly chosen as the training images while the remaining images were used as the test images. The classification of these test images were already known, which could be used as the ground truth to calculate the classification accuracy. We varied the values of k and n to change the training data size.

Experiments on image classification: For each set of k and n , image classification experiments were done using ADM, BDM, and bagging, respectively. Table 1 shows the result. For all the different values of k and n , our method achieved the highest classification accuracy. Note that although these accuracy values are lower than those reported in [1], a significantly smaller number of training images was employed in this study.

Preliminary experiments on image retrieval: ADM and BDM were also compared in the context of image retrieval. We randomly selected test image as the query image to retrieve images in the training data. Figure 1 shows a few examples of the retrieval when $k = 6$ and $n = 3$, where only the 3 closest matches found by ADM and BDM are shown. Our method gave a better result as the retrieved images contains fewer irrelevant results.

4.2 Experiments Using Non-clinical Images

The non-clinical images used are the toy car images², which consists of 255 images of 14 classes. Like experiments using clinical images, the values of k and n are varied to change the size of training data.

Experiments on image classification: ADM was compared with BDM and bagging for each set of k and n in the same manner as for the clinical images. All color images were turned into grayscale before the experiments. Table 2

¹ http://ganymed.imib.rwth-aachen.de/irma/datasets_en.php

² <http://lear.inrialpes.fr/people/nowak/dwl/toycarlear.tar.gz>

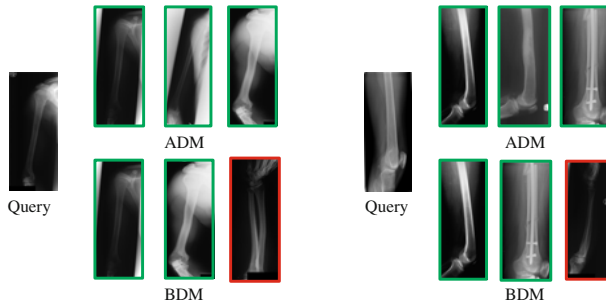


Fig. 1. Examples of clinical image retrieval result. Green and red bounding boxes indicate relevant and irrelevant retrieval result, respectively. Retrieved images are sorted according to their similarity to the query.

Table 2. Classification accuracy of non-clinical images using different methods. The highest accuracy values are shown bolded.

Number of classes (k)	Number of training image per class (n)	Number of test images	Classification accuracy		
			ADM	BDM	Bagging
6	3	94	0.86	0.78	0.79
10	3	160	0.64	0.51	0.46
14	4	200	0.61	0.46	0.41
10	6	130	0.69	0.66	0.67

shows the comparison result. For all the different values of k and n , our method returned fewer irrelevant images.

Preliminary experiments on image retrieval: We also compared ADM and BDM in the context of image retrieval. Query image was randomly chosen to retrieve the training images. Figure 2 shows a few examples of the retrieval when $k = 6$ and $n = 3$, where only the 3 closest matches found by ADM and BDM are shown. Our method again gave a better result.



Fig. 2. Two examples of non-clinical image retrieval result. Green and red bounding boxes indicate relevant and irrelevant retrieval result, respectively. Retrieved images are sorted according to their similarity to the query.

4.3 Discussion

We found that our method performs better than bagging when the training data for each class are limited. We suspect the reason is that the plurality vote, which is employed by bagging, sometimes results in two or more classes tied in a vote. In such cases, bagging selects a class randomly. However, when evaluated by the summation of numeric distances, as in our method, it is less likely to result in a tie.

In Section 3.2, at each iteration, we randomly choose n whole classes of images to construct the sub-sets. We would like to point out that theoretically, there should be no restriction on how the sub-sets are constructed. The reason behind our choice is that all the training images of a certain class can be utilized. Since the number of training images for each class is already small, further reducing this number may run the risk of decreasing the statistical representativeness of the data even more [4]. Also, in this way, we manage to place a reasonable upper limit on the total number of sub-sets.

5 Concluding Remarks

Accurate image classification is important in many medical applications. Supervised distance metric learning has shown great potential in image classification tasks. However, when faced with limited training data, especially when the number of training images for each class is small, the classification accuracy may be severely affected. We presented a novel method called *aggregated distance metric learning* (ADM) to classify a test image with limited number of training images. Our method combines the best of distance metric learning and bagging. Our method was found to be more accurate compared to both BDM and the bagging approach. We also presented a rigorous theoretical analysis to demonstrate that ADM is better at image classification compared to BDM. Experimental results using both clinical and non-clinical image data showed the efficacy of our method. Future work will involve further testing of our method on additional data sets and performing more extensive quantitative evaluation.

References

1. Yang, L., Jin, R., Mummert, L., Sukthankar, R., Goode, A., Zheng, B., Hoi, S., Satyanarayanan, M.: A boosting framework for visuality-preserving distance metric learning and its application to medical image retrieval. *IEEE PAMI* 32, 30–44 (2010)
2. Xing, E., Ng, A., Jordan, M., Russell, S.: Distance metric learning, with application to clustering with side-information. In: *NIPS*, pp. 521–528 (2002)
3. Hertz, T., Hillel, A., Weinshall, D.: Learning a kernel function for classification with small training samples. In: *ICML*, pp. 401–408 (2006)
4. Dekel, O., Shamir, O.: Multiclass-multilabel classification with more classes than examples. In: *International Conference on Artificial Intelligence and Statistics*, pp.137–144 (2010)
5. Zhang, T.: Class-size independent generalization analysis of some discriminative multi-category classification. In: *NIPS*, pp. 1625–1632 (2004)
6. Breiman, L.: Bagging predictors. *Machine Learning* 24, 123–140 (1996)

Sparse Classification for Computer Aided Diagnosis Using Learned Dictionaries

Meizhu Liu, Le Lu, Xiaojing Ye, Shipeng Yu, and Marcos Salganicoff

Siemens Medical Solutions USA, Malvern, PA 19355

le-lu@siemens.com

University of Florida, Gainesville, FL 32611

mliu@cise.ufl.edu

Abstract. Classification is one of the core problems in computer-aided cancer diagnosis (CAD) via medical image interpretation. High detection sensitivity with reasonably low false positive (FP) rate is essential for any CAD system to be accepted as a valuable or even indispensable tool in radiologists' workflow. In this paper, we propose a novel classification framework based on sparse representation. It first builds an overcomplete dictionary of atoms for each class via K -SVD learning, then classification is formulated as sparse coding which can be solved efficiently. This representation naturally generalizes for both binary and multiwise classification problems, and can be used as a standalone classifier or integrated with an existing decision system. Our method is extensively validated in CAD systems for both colorectal polyp and lung nodule detection, using hospital scale, multi-site clinical datasets. The results show that we achieve superior classification performance than existing state-of-the-arts, using support vector machine (SVM) and its variants [1, 2], boosting [3], logistic regression [4], relevance vector machine (RVM) [5, 6], or k -nearest neighbor (KNN) [7].

1 Introduction

Colon cancer and lung cancer are the two leading causes of cancer deaths in western population. However, these two cancers are highly preventable or "curable" if detected early. Image interpretation based cancer detection via 3D computer tomography has emerged as a common clinical practice, and many computer-aided detection tools for enhancing radiologists' diagnostic performance and effectiveness are developed in the last decade [1-4, 6-8]. The key for radiologists to accept the clinical usage of a CAD system is highest possible detection sensitivity with reasonably low false positive (FP) rate per case.

CAD system generally contains two stages: *Image Processing* as extracting (sub)volumes of interest (VOI) by heuristic volume parsing, and informative feature attributes describing the underlying (cancerous) anatomic structures; *Classification* as deciding the class assignment (cancer, or non-cancer) for selected VOIs by analyzing features. VOI selection is also called *candidate generation*, or CG to rapidly identify possibly anomalous regions with high sensitivity, but low

specificity, e.g. more than 100 candidates per scan with one to two true positives. Then dozens or hundreds of heterogeneous image features can be computed per VOI, in domains of volumetric shape, intensity, gradient, texture and even context [1,3,4,7,8]. Lastly, the essential goal for classification is achieving the best balance between high sensitivities and low false positive rates, given VOIs and associated features.

In this paper, we propose a new sparsity conducted classification framework, namely *dictionary learning as training* and *sparse coding as testing*, for CAD problems. Sparse signal representation has proved to be a very powerful tool for robustly acquiring, representing, and compressing high-dimensional signals that can be accurately constructed from a compact, fixed set of basis. The sparse representation (related to but different from subspace models of principal component analysis, independent component analysis, non-negative matrix decomposition) is effective in pattern recognition problems, and with link to biological evidence in human cortex system [9]. To the best of our knowledge, the present paper is the first reported work of exploiting sparse representation for CAD classification.

Different from the conventional *parametric* supervised classifiers of SVM, RVM, KNN, logistic regression and so on, a *nonparametric vocabulary* as set of exemplary atoms (learned as optimal rank-1 data matrix approximations) is constructed by maximizing its reconstruction power (or minimizing reconstruction error), within each positive or negative class, given the original training dataset. Then the testing or classification of a new data sample is accomplished by solving for the best approximation per vocabulary/class, under various sparsity constraints. The proposed classification method is evaluated on two large scale clinical datasets collected from multiple clinical sites across continents, for two tasks of colon polyp and lung nodule detections. Our datasets are representative, but very challenging with large within-class variations for polyp, nodule class and other anatomical structures in colon and lung volumes. The results validate that this new classification framework can significantly improve the accuracy of our baseline computer-aided detection system, using the same set of input image features, and compare favorably with other state-of-the-arts [1-4,6-8].

2 A Dictionary Approach to Classification

In this section, we present the new sparsity based classification framework for both binary or multiwise classes. The framework is comprised of two steps: dictionary learning and sparse coding. Unless otherwise noted, all vectors in this paper are column vectors. Also, $\|\cdot\|_2$ represents the regular Euclidean norm, and $\|\cdot\|_0$ counts the number of nonzero components of a vector. $(;\cdot)$ denotes a vector or matrix by stacking the arguments vertically.

2.1 Sparse Dictionary Learning

Problem Formulation: Suppose that there are N data samples $\{y_i \in \mathbb{R}^n : i = 1, \dots, N\}$ of dimension n , and the collection of these N samples forms an n -by- N

data matrix $Y = (y_1, \dots, y_N)$ with each column as one sample vector. Our goal is to construct a representative dictionary for Y , in the form of an n -by- K matrix $D = (d_1, \dots, d_K)$, that consists of K (usually $K \ll N$) key features $\{d_i \in \mathbb{R}^n : i = 1, \dots, K\}$ extracted from Y . In the dictionary context, d_i is also called an atom that represents one prototype feature in the category. This dictionary D needs to be trained from Y , and should be capable to sparsely represent *all* the samples that are in the same category as those in Y . Here by sparse representation we mean that each y_i can be written as a linear combination of very few atoms in D . In other words, we want to find a dictionary D and corresponding coefficient matrix $X = (x_1, \dots, x_N) \in \mathbb{R}^{K \times N}$ such that $y_i = Dx_i$ and $\|x_i\|_0 \ll K$ for all $i = 1, \dots, N$.

The problem can be readily formulated as the following minimizations:

$$\min_{D, X} \sum_{i=1}^N \|x_i\|_0, \text{ subject to } \|y_i - Dx_i\|_2 \leq \epsilon, \quad i = 1, \dots, N, \quad (1)$$

where $\epsilon > 0$ is the prescribed error tolerance of representation error. The solution (D, X) of (1) yields a dictionary D which extracts the main features $\{d_k : k = 1, \dots, K\}$ from samples in Y , and a coefficient matrix X with each column x_i representing the correlations between y_i and the dictionary atoms in D , by $\min_{D, X} \|x_i\|_0$.

Solving D, X : Since the objective function in (1) is highly nonconvex and nonsmooth, the solution is in general nontrivial. However, there are several algorithms that can be used to well approximate the solutions of (1), and numerous numerical tests demonstrated that these algorithms are very effective in practice. In this paper, we use the recently developed K -SVD algorithm [10], which has proved to be very robust to solve (1), by iterating exact K times of Singular Value Decomposition (SVD).

Starting from an initial dictionary, K -SVD algorithm approaches the solution of (1) by alternating the following two steps: the minimization with respect to X with D fixed, and the update of atoms in D using the current X .

The first step is called the ‘‘sparse coding’’ and can be formulated as

$$\min_{x_i} \|x_i\|_0, \text{ subject to } \|y_i - Dx_i\|_2 \leq \epsilon, \quad i = 1, \dots, N, \quad (2)$$

Although (2) is in general an NP-hard problem, the solution can usually be well approximated by many pursuit algorithms. In this work, we used the default sparse coding solver in K -SVD algorithm called the orthogonal matching pursuit (OMP) [11].

The second step is called the ‘‘dictionary update’’ which modifies the atoms in D one by one to better represent the data Y . To update d_k , the K -SVD algorithm first finds the index set $I_k = \{i : x_{ki} \neq 0\}$, which is just the set of indices of y_i 's who used d_k in representation in the sparse coding step. Then it applies the singular value decomposition (SVD) of the error matrix

$$E_k = Y_k - D_k X_k \quad (3)$$

where D_k is D with d_k replaced by 0. In (3), Y_k and X_k collect the columns with indices in I_k from Y and X , respectively. Finally, K -SVD substitutes d_k in D by the principal singular vector from the SVD of E_k and modifies the coefficients accordingly. This optimization is sequentially executed for each $k = 1, \dots, K$ while keeping all other columns d_j ($j \neq k$) fixed. Refer [10] for more details.

Output: The output of K -SVD consists of a trained dictionary D that contains atoms as features extracted from Y , and a coefficients X that records the sparse correlation or dependency of each sample y_i to these atoms. This learned dictionary D will be employed as a special form of classifier for CAD classification task. The dictionary D is build from training data Y in a data driven manner, and is capable to sparsely represent the very majority of data samples that are similar to those in Y .

2.2 Classification Using Learned Dictionaries

Our sparsity based classification framework, including the dictionary learning and classifier building, is essentially *generative*. It is able to handle both binary and multiwise classification problems.

Suppose that the training samples are given in the form of L ($L \geq 2$) categories, $\{Y^{(l)} \in \mathbb{R}^{n \times N_l} : l = 1, \dots, L\}$, where $Y^{(l)} = (y_1^{(l)}, \dots, y_{N_l}^{(l)})$ consists of N_l training samples labeled by l . To design a robust classifier, we apply the K -SVD algorithm to (1) with $Y = Y^{(l)}$, and obtain the respective dictionary $D^{(l)}$ for each $l = 1, \dots, L$. Now $D^{(l)}$ consists of the main exemplary atoms or features of the l -th category, and all samples belonging to this category can be sparsely represented by $D^{(l)}$. Furthermore, we can construct the global dictionary D by concatenating all $D^{(l)}$ as follows

$$D = (D^{(1)}, D^{(2)}, \dots, D^{(L)}) \in \mathbb{R}^{n \times N} \quad (4)$$

where $N = \sum_l N_l$. This global dictionary D is used as the classifier in our tests.

In order to determine the label of a new coming sample y , we solve the minimization problem

$$\min_x \|x\|_0, \text{ subject to } \|y - Dx\|_2 \leq \epsilon \quad (5)$$

with the global dictionary D in (4) using OMP [11]. Then we can examine the coefficient vector $x = (x^{(1)}; \dots; x^{(L)})$ solved from (5), and classify y to the l -th category if the nonzero components of x are clustered in the l -th segment of x , i.e. $x^{(l)}$. That is, a label l is assigned to y if the solution $x = (x^{(1)}; \dots; x^{(L)})$ of (5) satisfies

$$\|x^{(l)}\|_0 = \max\{\|x^{(m)}\|_0 : m = 1, \dots, L\}. \quad (6)$$

An ambiguous situation may happen if $x^{(l')}$ contains the largest component of x , but has less nonzero elements when compared to $x^{(l)}$. In this case, the label assigned to y is l instead of l' according to the criterion (6). However, it is more intuitive and reasonable to assign y by the label l' , as the key feature of

y occurs with higher weights in the l' -th category or $D^{(l')}$. A remedy of this is to substitute the ℓ_0 norm in (6) by the ℓ_1 -norm which can retain the sparsity property and take the magnitudes of the coefficients into account.

An alternative criterion for classification is to solve the per-category objective

$$\min_{z^{(l)}} \|z^{(l)}\|_0, \text{ subject to } \|y - D^{(l)}z^{(l)}\|_2 \leq \epsilon \quad (7)$$

for $l = 1, \dots, L$ respectively, and obtain the coefficients $z^{(l)} \in \mathbb{R}^{N_l}$ for all $l = 1, \dots, L$. Then y is classified to the l -th category if y appears to be “more” sparse with respect to $D^{(l)}$, namely,

$$\|z^{(l)}\|_0 = \min\{\|z^{(m)}\|_0 : m = 1, \dots, L\}. \quad (8)$$

This means that $D^{(l)}$ is more capable to extract the key features, or components of y than other dictionaries, indicating that y should be in the category l .

It is worth noting that the difference between (5) and (7) leads to distinct criterion (6) and (8). The criterion (6) implies that y is more similar to the contents in $D^{(l)}$ so it prefers $D^{(l)}$ when exposed to all $D = \{D^{(l)}\}$ simultaneously. On the other hand, (8) suggests that $D^{(l)}$ effectively attains the main features of y and is more capable to represent y sparsely when compared to other dictionaries, with the same error tolerance ϵ .

3 Experiments

Data: Our colon CAD dataset contains 429 patients or 858 CT volumes (i.e., two prone/supine scans per patient), collected from multiple hospitals in the US, Canada, Asia and Europe, and acquired using Siemens, GE and Philips scanners. After the candidate generation process (briefly discussed in Section 2), we obtain 134116 data candidates, out of which 1116 samples are positives belonging to 391 real polyps because one polyp can have multiple instances appeared, and the rests are negatives. Each data sample is represented using a 96 dimensional feature vector, including geometry, shape morphology, intensity and texture cues, computed by our CAD system. Moreover, 411 positive samples are instances of 137 **flat** polyps and 705 positives belong to 254 **non-flat**, or **SP** (e.g., sessile, pedunculated and mass) polyps. Therefore, the dataset can be subdivided as two classes as negatives (-) and positives (+) or three categories, namely negatives (-1), flat polyps (+1) and non-flat polyps (+2). The lung nodule dataset was obtained from 1000 patients from multiple medical sites in different countries using various scanners. This dataset contains the information of part-solid nodules with a diameter range of 4-20mm. There are 49,094 samples after CG stage, out of which 2,531 are positive nodule instances (+) and the rest as negatives (-). Each data sample has 112 features. In the following, our experiments use 5 fold cross-validation and no data samples from the same patient are used for both training and testing.

Standalone: For comparison, we first train a baseline classifier using multiple instance relevance vector machine (MILRVM) [5], and its training/testing classification performances, in the form of Free-Response Operator Characteristic

(FROC) curves, are illustrated in Fig. 1 and 2. Note that these baselines achieve comparable results with state-of-the-arts [1-4, 6-8] on datasets of similar data scales). Next, using the positive and negative samples in the training dataset (i.e., $L = 2$), we learned dictionaries $\mathbf{D}_+ \in \mathbb{R}^{n \times K}$ and $\mathbf{D}_- \in \mathbb{R}^{n \times K}$ for (+/-) classes, respectively. The dictionary size $K = \beta \times n$ is normally chosen with respect to data dimension n , here $\beta = 4$. When $\beta > 4$, the classification performances are similar. After this, both dictionaries are concatenated into a single dictionary $\mathbf{D} \in \mathbb{R}^{n \times (L \times K)}$, and the classification criterion (5) is applied. In colon dataset $n = 96$ and $K = 384$; and $n = 112$, $K = 448$ for lung. The sensitivities are calculated on per-polyp or per-nodule level, consist with multiple instance learning setting [5], and the false positive (FP) rates are reported on a per-patient level (i.e., summing FPs in two volume views) in colon and per-volume level for lung. For $L = 3$ of colon CAD, when an instance is classified either as (+1) for flats or (+2) for SP, an overall ‘‘hit’’ or detection will be counted. The classification results on training and testing (validation) datasets are shown in Table 1. Though sparse classification does not provide FROC curves, it has (2% ~ 5%) higher detection sensitivities than our MILRVM baselines, at low FP rates of ≤ 3 per case in both colon and lung datasets. This is highly suitable for clinical applications.

Table 1. Standalone Sparsity Classification Results for Colon Polyp (L=3) and Lung Nodule Detection (L=2)

	Colon Polyp CAD				Lung Nodule CAD	
	FP Rate	Sensitivity	Flat Sensitivity	SP Sensitivity	FP Rate	Sensitivity
Training	2.6818	91.12%	84.97%	91.79%	2.6919	90.32%
Testing	2.6897	89.68%	79.98%	94.20%	2.6797	89.65%

Gated Fusion: To build the best overall CAD system, we exploit the *three-way gated decision tree*, integrating both RVMMIL [5] and sparsity classifiers. RVMMIL assigns each data sample a probability value $\rho(+)$, of being positive class. Therefore, we design the following three gates or decision rules: (1) if $\rho(+)$ $\geq \gamma_1$, classifying as positive; (2) if $\rho(+)$ $\leq \gamma_2$, classifying as negative, where $\gamma_1 > \gamma_2$; (3) if $\gamma_1 > \rho(+)$ $> \gamma_2$, employing sparse classification (L=2, or L=3). The thresholds γ_1, γ_2 are estimated by maximizing the decision tree classification accuracy via cross validation. Conditions $\rho(+)$ $\geq \gamma_1$, $\rho(+)$ $\leq \gamma_2$ indicate samples being positive or negative *with high confidence*; while $\gamma_1 > \rho(+)$ $> \gamma_2$ refers *ambiguous* classifying data samples by RVMMIL.

Fig. 1 (Left) shows the combined model achieves 6% ~ 8%, or 2.4% ~ 3.2% sensitivity improvements for colon polyp detection, in training and testing respectively, at ~ 2.7 FPs per patient. On the other hand, at the same sensitivities, our method can reduce the FP rates by 3 ~ 4 per patient, with respect to training or testing. Note that $L = 3$ performs better than $L = 2$ which shows the advantage of modeling capacity for a more comprehensively generative representation, consisting of richer dictionaries $D = (D^{(1)}, D^{(2)}, \dots, D^{(L)})$. Similarly, at least

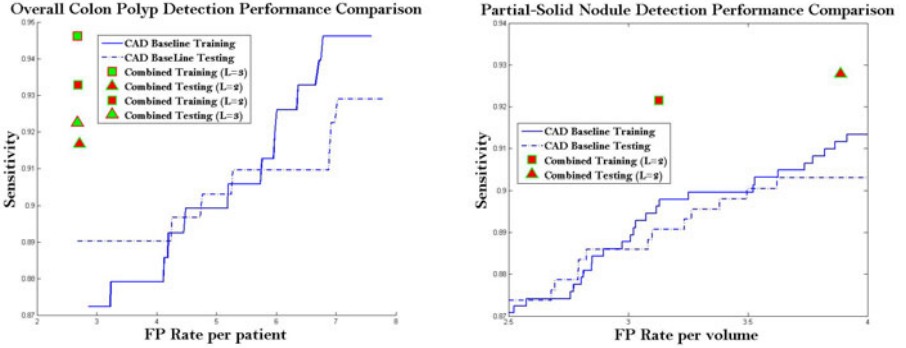


Fig. 1. The classification results of using our proposed method and comparison with the CAD baseline, for training and testing in the colon dataset (**Left**) and lung dataset (**Right**). CAD baselines are plotted for comparison.

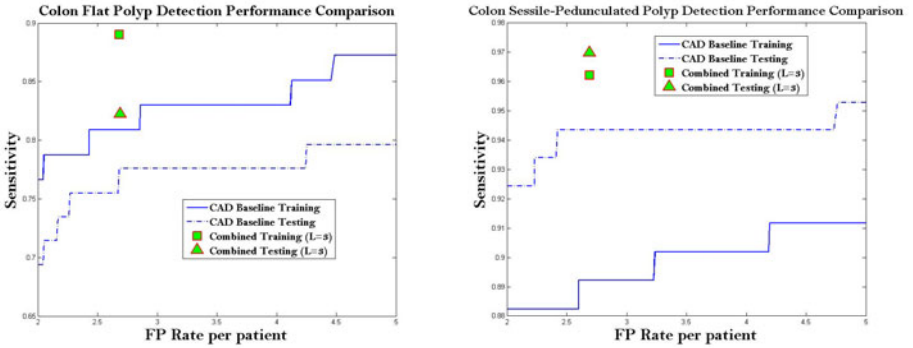


Fig. 2. Sensitivity vs. FP rate per patient (i.e., two volumes) for Flat polyp detection (**Left**) in training and testing datasets; and for Sessile-Pedunculated polyp detection (**Right**). CAD baselines are plotted for comparison.

2% ~ 3% sensitivity gains are observed for lung nodule detection, corresponding to the same FP rates, in Fig. 1 (Right). These improvements are statistically significant for colon/lung cancer detections, from the already high-performed baselines. Furthermore, the sensitivities for flat or sessile-pedunculated polyps are also greatly increased (by 4% ~ 7%), as shown in Fig. 2 (**Left**) and (**Right**). The three-gate combined model also consistently outperforms single RVMMIL baseline or sparsity classifier, in all scenarios.

4 Conclusion and Future Work

In this paper, we present a new sparse representation based classification method for computer-aided diagnosis problem, by learning an overcomplete dictionary of

exemplary atoms for each data class and adapting sparse coding criteria for effective classification. This generative formulation has the ability of modeling two or multiple classes in the same way. It can be used either as a standalone classifier, or integrated with other decision-making scheme(s). Our proposed method is validated in two CAD systems of colorectal polyp and lung nodule detection, using large scale, representative clinical datasets. The results show that we achieve superior performances than our baseline and other existing state-of-the-arts. In future work, we plan to explore how to integrate class discriminative information for dictionary learning [12], and other decision fusion structures of heterogeneous classifiers.

References

1. Wang, S., Yao, J., Summers, R.: Improved Classifier for Computer-aided Polyp Detection in CT Colonography by Nonlinear Dimensionality Reduction. *Medical Physics* 35, 1377–1386 (2008)
2. Bogoni, L., Bi, J.: Lung Nodule Detection. *ImageCLEF: Experimental Evaluation in Visual Information Retrieval* 32, 415–434 (2010)
3. Slabaugh, G., Yang, X., Ye, X., Boyes, R., Beddoe, G.: A Robust and Fast System for CTC Computer-Aided Detection of Colorectal Lesions. *Algorithms* 3(1), 21–43 (2010)
4. van Ravesteijn, V., van Wijk, C., Vos, F., Truyen, R., Peters, J., Stoker, J., van Vliet, L.: Computer Aided Detection of Polyps in CT Colonography using Logistic Regression. *IEEE Trans. on Med. Imag.* (2010)
5. Raykar, V., Krishnapuram, B., Bi, J., Dundar, M., Rao, R.: Bayesian Multiple Instance Learning: Automatic Feature Selection and Inductive Transfer. In: *ICML*, pp. 808–815 (2008)
6. Mang, T., Bogoni, L., Salganicoff, M., Wolf, M.: CT Colonography: Retrospective Evaluation of the Performance of Computer-aided Detection of Colonic Polyps in Tagged and Untagged Preparation. In: *ECR* (2010)
7. van Ginneken, B., Schilhama, A., Prokopy, M.: A large-scale evaluation of automatic pulmonary nodule detection in chest CT using local image features and k-nearest-neighbour classification. *Med. Image Anal.*, 670–70 (2009)
8. van Ginneken, B.: Comparing and Combining Algorithms for Computer-aided Detection of Pulmonary Nodules in Computed Tomography Scans: The ANODE09 Study. *Med. Image Anal.*, 707–722 (2010)
9. Vinje, W., Gallant, J.: Sparse coding and decorrelation in primary visual cortex during natural vision. *Science*, 1273–1276 (2000)
10. Aharon, M., Elad, M., Bruckstein, A., Katz, Y.: K-SVD: An Algorithm for Designing of Overcomplete Dictionaries for Sparse Representation. *IEEE Transactions on Signal Processing* 54, 4311–4322 (2006)
11. Tropp, J.: Greed is good: Algorithmic results for sparse approximation. *IEEE Trans. Inf. Theory* 50, 2231–2242 (2004)
12. Mairal, J., Bach, F., Ponce, J., Sapiro, G., Zisserman, A.: Learning discriminative dictionaries for local image analysis. In: *Proc. IEEE CVPR* (2008)

A Discriminative-Generative Model for Detecting Intravenous Contrast in CT Images

Antonio Criminisi¹, Krishna Juluru², and Sayan Pathak³

¹ Microsoft Research Ltd., CB3 0FB, Cambridge, UK

² Weill Cornell Medical College, New York, NY

³ Microsoft Corporation, One Microsoft Way, Redmond, US

Abstract. This paper presents an algorithm for the automatic detection of intravenous contrast in CT scans. This is useful e.g. for quality control, given the unreliability of the existing DICOM contrast metadata.

The algorithm is based on a hybrid discriminative-generative probabilistic model. A discriminative detector localizes enhancing regions of interest in the scan. Then a generative classifier optimally fuses evidence gathered from those regions into an efficient, probabilistic prediction.

The main contribution is in the generative part. It assigns optimal weights to the detected organs based on their learned degree of enhancement under contrast material. The model is robust with respect to missing organs, patients geometry, pathology and settings. Validation is performed on a database of 400 highly variable patients CT scans. Results indicate detection accuracy greater than 91% at ~ 1 second per scan.

1 Introduction

Medical images stored in the DICOM standard contain a wealth of associated metadata. Many metadata elements are acquired automatically, e.g. the slice thickness. Others (e.g. the anatomical region being scanned) require human input at the time of acquisition. However, manual input is prone to error [1]. Accuracy of DICOM metadata is important in routine clinical care, where those tags are used to ensure that the requested imaging service has in fact been performed. Metadata accuracy is also crucial in clinical research, e.g. in trials where the effectiveness of a drug is monitored using imaging.

This paper presents an algorithm for the automatic detection of *intravenous contrast*. Its corresponding DICOM tag (id (0018,0010), one of the key metadata) requires manual input during image acquisition. However, in our experience only $\sim 80\%$ of CT studies present the correct value (see also [1]). Such accuracy is operator and hospital dependent. The unreliability of this tag means that the only way to know whether there is contrast in a scan is through visual inspection.

Our technique alleviates this problem by automatically classifying CT scans as contrast-enhanced or not. In [2] Prince et al looked at detection of contrast material arrival in the aorta. However, manual intervention was required to select

regions of interest. Similarly in [3]. Such work would benefit greatly from our technique as it removes the need for any manual input.

Effective automatic detection of contrast material does require some level of localization of relevant regions (*e.g.* organs which enhance in the presence of i.v. contrast agents). Thus, techniques for anatomy segmentation in CT could be beneficial for such purposes. Many such algorithms have been proposed, *e.g.* [4,5,6,7,8]. However, the majority tend to focus on a single area of the body, or are specific to a certain disease and sometimes still rely on human intervention. Also, none of these have been used for verification of DICOM metadata accuracy. The algorithm proposed here automatically detects multiple enhancing regions of interests and combines the gathered evidence *optimally* to produce accurate contrast detection. It works on completely generic, pathology-rich CT scans.

Outline. Our algorithm is in two stages: I) first, major anatomical structures are automatically localized; II) then, contrast detection is achieved via a generative model of CT scans. The main novelty is in the second part which optimally fuses evidence from detected organs for robust contrast detection.

2 The Annotated Database

We have collected a database of 400 diverse patients CT scans (see fig. 1). The patients suffer from various pathologies and the scans have been acquired with different scanners, at varying resolutions, viewing different body regions. The data comes from hospitals in different continents. All studies are expert labeled into belonging to the “contrast-enhanced” (C) or “native” (N) classes. If only oral contrast is present (and no i.v. contrast) then the scan is labeled as N.

As detailed in section 5 the database is divided into many non-overlapping training and test sets. Parameters are optimized on the training set and all accuracy measurements performed on the previously unseen test set, to avoid over-fitting.

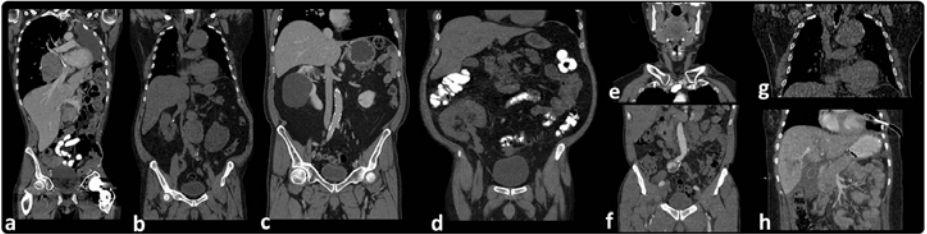


Fig. 1. Variability in our 400-patient database. (a, c, e, f, h) Patient scans with i.v. contrast. (b, d, g) No i.v. contrast. (a, d) Additional oral contrast. (a) Scan with missing left lung. (c) Kidney cyst. (d) Transplanted kidney. (f) Aortic aneurism.

3 Stage I - Discriminative Anatomy Localization

Our algorithm for the automatic localization of anatomy in CT builds upon the work in [9] and is summarized here for completeness. A regression forest is used to map CT voxels *directly* to organs position and extent. We build upon a forest technique because of its efficiency and robustness.

The forest is trained to localize the following 12 structures: **heart, liver, spleen, lungs, kidneys, gall bladder, pelvis, femurs**. The trees are trained on a set of volumes with known organ bounding boxes. The trees cluster voxels together based on their appearance, their spatial context and their confidence in predicting position and size of all anatomical structures.

During testing, all voxels are pushed into all trees. Each node applies a test on each voxel which consequently is sent to the left or right child. When a voxel reaches a leaf node the stored relative displacements are used to cast a vote on position and extent of each organ. The leaves with largest prediction confidence carry more weight in the final prediction. See [9] for further details.

Figure 2 shows some detection results. Good anatomy localization is achieved even in the presence of pathologies such as cysts or missing lungs. Our C++ implementation produces a mean localization error of $\sim 1.5cm$ in approximately $1s$ on a single-core desktop machine. Next we describe how information coming from all detected organs is fused together to yield accurate contrast detection.

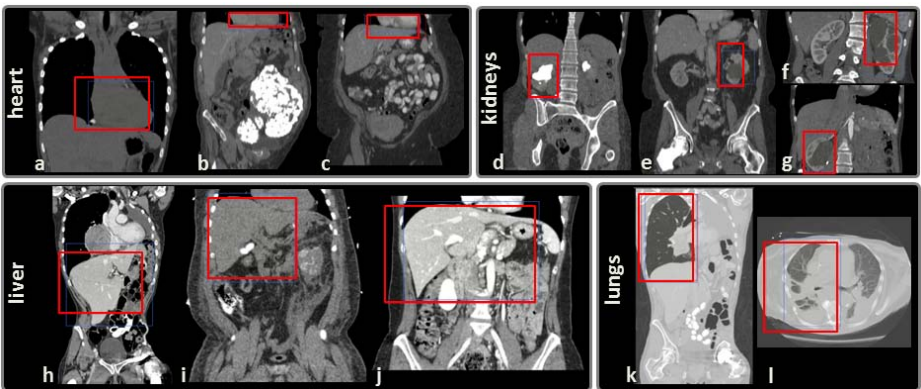


Fig. 2. Results of discriminative anatomy localization. Automatic anatomy detection in diverse patients' CT scans. The detected organs (*e.g.* heart, kidney, liver, lung) are shown in red. The faint blue line denotes ground truth. The detections are robust to large variabilities in the scans, including pathologies. *E.g.* (b) Large cyst in abdomen. (h, k) Missing left lung. (d, e, f, g) Cysts and other kidney anomalies.

4 Stage II - Generative Contrast/No-Contrast Classification

This section describes the main contribution of this paper, *i.e.* a new, part-based generative model of CT scans and its use for contrast detection. Given a previously unseen 3D CT image, the goal is that of estimating its class c , with $c \in \{\mathbf{C}, \mathbf{N}\}$. Here we also assume given the position of organs of interest together with the associated confidence (as described in section 3).

The probabilistic model. Evidence coming from detected anatomical regions is aggregated via the graphical model shown in fig. 3. The probability of class c is what we wish to estimate. The variable \mathbf{s}_i denotes the appearance of the i^{th} anatomical structure. In a given image, the organ appearance \mathbf{s}_i is represented as a histogram over density values (in HU). The histogram is computed over all voxels within the organ’s bounding box and is normalized to have unit area.

Handling visibility. In a given scan some organs may be absent (*e.g.* because missing or outside the capture area). Their status is captured by the binary visibility variable $o_i \in \{0, 1\}$, with 0, 1 denoting absence/presence, respectively. For the organ i we set $o_i = 0$ if the organ detection posterior falls below 0.5, as in 9. From fig. 3 the joint probability of the whole model is

$$p(c, \mathbf{S}, \mathbf{O}) = p(c) \prod_i^N p(\mathbf{s}_i | o_i, c) p(o_i) \quad (1)$$

with $N = 12$ the total number of organs, $\mathbf{S} = \{\mathbf{s}_1, \dots, \mathbf{s}_N\}$ and $\mathbf{O} = \{o_1, \dots, o_N\}$.

Model training. The data likelihood $p(\mathbf{s}_i | o_i = 1, c)$ is modeled as a multivariate Gaussian: $p(\mathbf{s}_i | o_i = 1, c) = \mathcal{N}(\mathbf{s}_i; \bar{\mathbf{s}}_i^c, \Lambda_i^c)$, with $\bar{\mathbf{s}}_i^c$ the mean and Λ_i^c the covariance matrix. Since c assumes binary values we need to learn two likelihood models for each organ. This is done readily from annotated training studies.

Figure 4 shows such likelihoods for some selected organ classes, learned from 200 training scans. We observe that the appearance of bony structures (*e.g.* the pelvis) is not affected by i.v. contrast. This is shown by the fact that $p(\mathbf{s} | c = \mathbf{C}) \approx p(\mathbf{s} | c = \mathbf{N})$ (visibility variable o removed here for simplicity). Instead, for organs

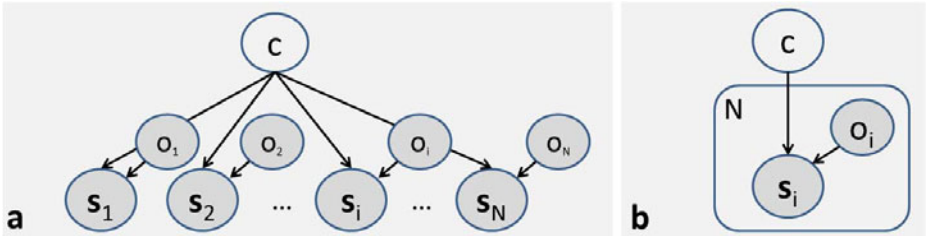


Fig. 3. Our generative model. (a) Graphical representation of our probabilistic model. (b) As in (a) but represented with plates. See text for details.

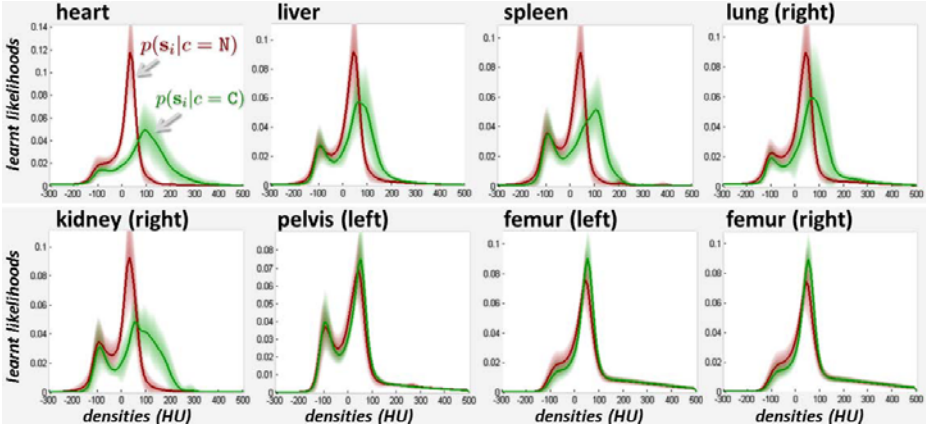


Fig. 4. Learned likelihood models. The learned likelihoods $p(\mathbf{s}_i | c = \mathbf{C})$ and $p(\mathbf{s}_i | c = \mathbf{N})$ for 8 organs are shown in green and red, respectively (visibility variables removed for simplicity). The two learned models are more different from one another in those organs that tend to enhance in the presence of i.v. contrast. Our algorithm assigns optimal weights (*e.g.* higher for heart and lower for femur) for final contrast detection.

such as the heart or kidneys (enhancing) the automatically learned likelihood models are very different in the two cases. This is not surprising, but what is important here is that our model captures such differences quantitatively. To obtain accurate contrast detection we wish to give more weight to those organs which enhance more in the presence of i.v. contrast, and vice-versa. But what are the right weights to use? Additionally, the appearance of some structures changes considerably across different patients while others show more consistency. Thus, uncertainty must be taken into account for the final contrast classification. Our generative model captures all these intuitions quantitatively and provides a way of assigning optimal weights to the different organs, automatically.

Contrast detection. The next step is to combine image observations and learned models and come up with a single contrast/no-contrast answer. During testing we are given a previously unseen scan, its detected bounding boxes, their associated appearance observations \mathbf{s}_i and also the visibility observations \mathbf{o}_i for all organs. Bayesian manipulation of (II) leads to the class posterior $p(c | \mathbf{S}, \mathbf{0}) = \frac{1}{Z} p(c) \prod_i^N p(\mathbf{s}_i | \mathbf{o}_i, c) p(\mathbf{o}_i)$, with $Z = \sum_{c=0}^1 \left[p(c) \prod_i^N p(\mathbf{s}_i | \mathbf{o}_i, c) p(\mathbf{o}_i) \right]$. So, the ratio of the class posteriors for the two cases is

$$R = \frac{p(c = \mathbf{C} | \mathbf{S}, \mathbf{0})}{p(c = \mathbf{N} | \mathbf{S}, \mathbf{0})} = \frac{p(c = \mathbf{C})}{p(c = \mathbf{N})} \prod_i^N \frac{p(\mathbf{s}_i | \mathbf{o}_i, c = \mathbf{C}) p(\mathbf{o}_i)}{p(\mathbf{s}_i | \mathbf{o}_i, c = \mathbf{N}) p(\mathbf{o}_i)} \quad (2)$$

with the priors $p(c = \mathbf{C})$ and $p(c = \mathbf{N})$ learned from the training set.

If a structure i is not present ($\mathbf{o}_i = 0$) then its appearance \mathbf{s}_i is undefined and so is the likelihood $p(\mathbf{s}_i | \mathbf{o}_i = 0, c)$. Thus we need to remove its influence

from (2). We do so by defining $p(\mathbf{s}_i|o_i, c) = p(\mathbf{s}_i|o_i = 1, c)^{o_i}$ which reduces to 1 for undetected structures. This when plugged into (2) has the desired effect. Thus the posterior ratio reduces to

$$R = \frac{p(c = \mathbf{C})}{p(c = \mathbf{N})} \prod_i^N \left(\frac{p(\mathbf{s}_i|o_i = 1, c = \mathbf{C})}{p(\mathbf{s}_i|o_i = 1, c = \mathbf{N})} \right)^{o_i} \quad (3)$$

Finally, maximum a-posteriori estimation (MAP) happens by checking the value of R in (3). If $R > 1$ then the scan is declared to be contrast-enhanced, and vice-versa. Values of R close to 1 indicate uncertain output.

In our model robustness to pathology arises from: i) having those anomalies in the training set, and ii) aggregating evidence over *multiple* organs. A small number of such organs may exhibit unusual shape or appearance, but other healthy structures would contribute positively towards the final outcome.

5 Experiments, Results and Comparisons

This section presents evaluations and comparisons with possible alternatives.

Figure 5 shows the learned likelihood models for various organs of interest and, for two test scans, the individual organ appearances \mathbf{s}_i . Note how organ appearance in contrast-enhanced scans is better explained by contrast-enhanced likelihood models, and vice-versa. This holds true for enhancing organs and does not for bony structures (whose black curve is equally close to both models).

When evaluating the posterior ratio (3) we see that each organ contributes to the final ratio via its own factor. All N factors then get multiplied together.

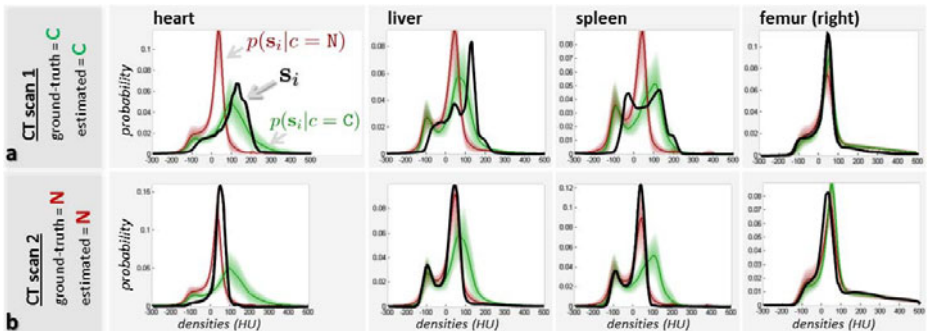


Fig. 5. Testing individual organ likelihoods for two previously unseen CT images. The learned likelihood models $p(\mathbf{s}_i|o_i = 1, c = \mathbf{C})$ and $p(\mathbf{s}_i|o_i = 1, c = \mathbf{N})$ for four different anatomical structures are shown in green and red, respectively. The appearance descriptor \mathbf{s}_i for the same organs in the input test scans are shown in black. (a) In the case of a contrast-enhanced image the organ appearances are better predicted by the contrasted likelihood models (black curve closer to the green models), (b) and vice-versa in the case of native CT scans. Such likelihoods are combined with prior probabilities in (3) to produce maximum a-posteriori classification.

Anatomical structures such as bony ones tend to contribute posterior factors close to 1 and thus they (correctly) do not have much effect on the final classification. On the contrary, enhancing soft tissue organs contribute in proportion to how much they enhance in the presence of i.v. contrast. This demonstrates automatic weighting of organs contribution for contrast detection.

Quantitative evaluation is done by repeatedly subdividing our 400-large CT database into disjoint training and testing sets, with random selection of the studies and varying ratio between the two sets. Our models are trained exclusively on the training set and evaluated exclusively on previously unseen test images. Figure 6a shows classification accuracy as a function of percentage of training volumes. As expected the accuracy increases with the number of training images. However, its behavior is very flat (close to constant), thus indicating good generalization, with a mean accuracy of $\sim 91\%$.

We also investigated the robustness of the generative classifier with respect to noise in the discriminative anatomy localizer. We do so by artificially adding varying amount of noise to the detected organs bounding boxes. Plotting accuracy as a function of noise (fig. 6b) shows a reasonably slow fall off for noise as large as $20mm$. This indicates good robustness with respect to inaccurate organ localization. In both plots in fig. 6 the experiments were repeated 100 times for each x-axis value. The estimated 2-std confidence region is shown shaded.

Visual inspection of incorrectly classified images shows that most of those are borderline cases, where perhaps the contrast agent has not had sufficient time to diffuse. Generally this uncertainty is indicated by a posterior ratio close to 1.

Efficiency. Run-times remain of the order of 1s for the whole discriminative + generative testing phase, as the generative part adds a negligible load.

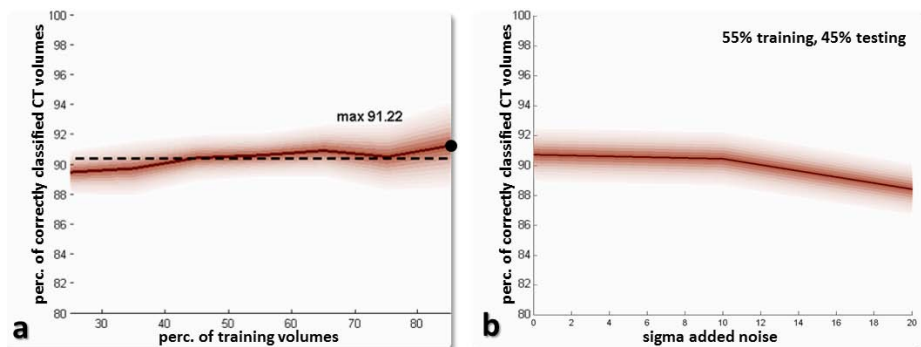


Fig. 6. Automatic contrast detection results (a) Results of scan classification into contrast-enhanced or native as a function of training/test ratio. Accuracy increases with the number of training data. However, the flatness of the curve indicates good generalization. (b) Results of contrast detection for increasingly noisy organ localization. The accuracy curve falls off slowly with increasing noise, thus indicating good noise robustness. Note that accuracy of existing manual DICOM tag is $\sim 80\%$.

Parameters. The only additional parameter in our generative model is the histogram length, which is automatically optimized during training.

Comparisons. We compared our algorithm against one of the state of the art recognition techniques, *i.e.* random classification forests [10]. Recently, classification forests have been shown to yield excellent results in diverse applications [11]. During training Haar wavelet-like features are automatically selected in different spatial arrangements and positions. This is to allow automatic alignment of box-like ROIs with enhancing organs and avoid a separate organ detection step. After optimizing all parameters for best results we recorded an accuracy of $\sim 87\%$, considerably lower than that of the two-step approach proposed here.

6 Conclusion

We have presented a hybrid discriminative-generative model for the efficient, automatic detection of intravenous contrast agent in CT scans. The algorithm can automatically localize enhancing regions in the scan and weigh them optimally in order to achieve the most accurate detection results. It is robust to missing organs, pathologies and possible inaccuracies in the organ localization stage.

Validation on a large database of pathology-rich scans has demonstrated accuracy greater than that of manually annotated DICOM contrast tags. Our algorithm plays an important role in routine quality control of hospital data as well as in clinical research. It may be applied to other quality control tasks such as detection of oral contrast or identification of the specific contrast material.

References

1. Guld, M., Kohnen, M., Keysers, D., Schubert, H., Wein, B., Bredno, J., Lehmann, T.: Quality of DICOM header information for image categorization. In: Proc. SPIE, San Diego (2002)
2. Prince, M., Chenevert, T., Foo, T., Londy, F., Ward, J., Maki, J.: Contrast-enhanced abdominal mr angiography: optimization of imaging delay time by automating the detection of contrast material arrival in the aorta. *Radiology (RSNA, Inc.)* 203 (April 1997)
3. Sheiman, R., Prassopoulos, P., Raptopoulos, V.: CT detection of layering of i.v. contrast material in the abdominal aorta. *American Journal of Roentgenology (American Roentgen Ray Society)* 171(5) (November 1998)
4. Bae, K., Kim, J.S., Na, Y.H., Kim, K., Kim, J.H.: Pulmonary nodules: automated detection on CT images with morphologic matching algorithm—preliminary results. *Radiology (RSNA, Inc.)* 236 (July 2005)
5. Nappi, J., Dachman, A., MacEneaney, P., Yoshida, H.: Automated knowledge-guided segmentation of colonic walls for computerized detection of polyps in CT colonography. *Journal of Computer Assisted Tomography* 26 (2002)
6. Gao, L., Heath, D., Kuszyk, B., Fishman, E.: Automatic liver segmentation technique for three-dimensional visualization of CT data. *Radiology (RSNA, Inc.)* 201 (November 1996)

7. Seifert, S., Barbu, A., Zhou, S.K., Liu, D., Feulner, J., Huber, M., Sühling, M., Cavallaro, A., Comaniciu, D.: Hierarchical parsing and semantic navigation of full body ct data. In: Pluim, J.P.W., Dawant, B.M. (eds.) SPIE (2009)
8. Soler, L., Delingette, H., Malandain, G., Montagnat, J., Ayache, N., Koehl, C., Dourthe, O., Malassagne, B., Smith, M., Mutter, D., Marescaux, J.: Fully automatic anatomical, pathological, and functional segmentation from ct scans for hepatic surgery. *Computer Aided Surgery: Official Journal of the International Society for Computer Aided Surgery* 6 (January 2001)
9. Criminisi, A., Shotton, J., Robertson, D., Konukoglu, E.: Regression forests for efficient anatomy detection and localization in CT studies. In: MICCAI Workshop on Medical Computer Vision: Recognition Techniques and Applications in Medical Imaging. Springer, Beijing (2010)
10. Breiman, L.: Random forests. Technical Report TR567, UC Berkeley (1999)
11. Shotton, J., Johnson, M., Cipolla, R.: Semantic texton forests for image categorization and segmentation. In: IEEE CVPR (2008)

Classification of Alzheimer’s Disease Using a Self-Smoothing Operator

Juan Eugenio Iglesias, Jiayan Jiang, Cheng-Yi Liu, Zhuowen Tu, and the
Alzheimers Disease Neuroimaging Initiative

Laboratory of Neuro Imaging, University of California, Los Angeles
{jeiglesias,chengyiliu}@ucla.edu, {jet.jiang,zhuowen.tu}@loni.ucla.edu

Abstract. In this study, we present a system for Alzheimer’s disease classification on the ADNI dataset [1]. Our system is able to learn/fuse registration-based (matching) and overlap-based similarity measures, which are enhanced using a self-smoothing operator (SSO). From a matrix of pair-wise affinities between data points, our system uses a diffusion process to output an enhanced matrix. The diffusion propagates the affinity mass along the intrinsic data space without the need to explicitly learn the manifold. Using the enhanced metric in nearest neighborhood classification, we show significantly improved accuracy for Alzheimer’s Disease over Diffusion Maps [2] and a popular metric learning approach [3]. State-of-the-art results are obtained in the classification of 120 brain MRIs from ADNI as normal, mild cognitive impairment, and Alzheimer’s.

1 Introduction

Alzheimers Disease (AD) and its preclinical stage, mild cognitive impairment (MCI), are the most common form of dementia in elders. Magnetic resonance imaging (MRI) can provide insight into the relation between AD and the structure of the brain: AD is known to be connected with gray matter loss [4] and with the shape of subcortical structures (especially the hippocampus) [5]. There have been several attempts in the literature to automatically classify a brain MRI as AD, MCI or normal (typically represented by older control subjects, OC). *Chupin et al.* [6] automatically segment the hippocampus and use its volume for the classification. *Vemuri et al.* [7] use support vector machines (SVM) based on tissue densities and a number of covariates (demographics, genotype). *Klöppel et al.* [8] feed a SVM directly with image data after registration (i.e. spatial alignment). *Zhang et al.* [9] use a SVM with cerebrospinal fluid, positron emission tomography and MRI data as features. *Davatzikos et al.* [10] use the distribution of gray matter, white matter and cerebrospinal fluid in registered space. *Desikan et al.* [11] feed the entorhinal cortex thickness, hippocampal volume and supramarginal gyrus thickness to a logistic regression analysis. These works are summarized in Table 1.

In this paper, we approach the OC/MCI/AD classification problem from the perspective of metric learning. Given a number of heterogeneous affinity measures between the data points, the task is to find an enhanced metric which will

Table 1. Representative methods in the literature of AD classification. For *Chupin et al.* we report the range of results in a number of two-class classification problems.

Method	# Subjects	Classes (prevalences)	Classification rate
<i>Chupin et al.</i> [6]	605	OC (24%), MCI (49%), AD(27%)	60-80%
<i>Vemuri et al.</i> [7]	100	OC (50%), AD (50%)	89%
<i>Klöppel et al.</i> [8]	68	OC (50%), AD (50%)	94%
<i>Zhang et al.</i> [9]	103	OC(50%), AD(50%)	93%
<i>Zhang et al.</i> [9]	150	OC(34%), MCI(66%)	76%
<i>Davatzikos et al.</i> [10]	30	OC (50%), MCI (50%)	90%
<i>Desikan et al.</i> [11]	151	OC (62%), MCI (38%)	90%

ultimately improve the classification rate in a k -nearest neighbor (k NN) framework. Popular distance metric learning methods [12,3], which are mostly supervised, learn a Mahalanobis distance parametrized by a positive semi-definite matrix. However, the performance gain is rather limited because a global linear transform does not suffice to discriminate the data. Nonlinear versions exist, but it is difficult to find a kernel that provides good results. Non-parametric manifold learning techniques such as Isomap [13] do not necessarily provide a better metric, which limits their use in classification. They also have the disadvantage that explicitly estimating the manifold can be difficult and time consuming. Their application to medical image analysis has also been limited [14].

Here we adopt an unsupervised metric learning algorithm: self-smoothing operator (SSO). SSO enhances an input pair-wise affinity matrix similar to a Gram matrix. A smoothing kernel is built from the matrix and used to iteratively propagate the affinity mass between strongly connected neighbors, following the structure of the manifold without having to compute it explicitly. The framework can accommodate semi-supervised learning (i.e. taking advantage of not only labeled but also unlabeled examples to build a classifier [15]): even if unlabeled examples cannot be used in the k NN classification, they can still be considered in the prior diffusion, often bridging gaps between points with the same label. A feature selection method is incorporated into the design of the affinity matrix to improve the results. We apply the proposed framework to the AD classification problem with registration-based and overlap-based similarity measures, comparing the results with metric learning [3] and Diffusion Maps [2].

2 Materials

Brain MRI from 120 subject Brain MRI scans from 120 subjects (age 76.7 ± 6.4 years) are used in this study. The subjects were randomly selected from the ADNI dataset [1] under two constraints: 1) the scans are from the same cross section (12 months after the start of the study); and 2) the three classes (OC,MCI,AD) and the two genders are equally represented. The scans were acquired with T_1 -weighted MPRAGE sequences, skull-stripped with BET [16] and fed to Brain-Parser [17] to automatically extract 56 cortical and subcortical structures.

3 Methods

3.1 Self-Smoothing Operator

SSO is closely related to the Diffusion Maps algorithm [2], which defines diffusion distances between data samples to improve an input pair-wise affinity matrix. It introduces a global diffusion distance metric over data samples. Given the transition kernel H (a row-wise-normalized version of the pair-wise affinity matrix), the diffusion distance between data samples x_i and x_j at step t is defined as:

$$d_t^2(i, j) = \|h_t(i, \cdot) - h_t(j, \cdot)\|_{1/\phi_0}^2 = \sum_k \frac{1}{\phi_0(k)} (h_t(i, k) - h_t(j, k))^2$$

where $h_t(i, \cdot)$ is the i -th row of H^t , and ϕ_0 is the equilibrium distribution.

Instead of using an alien notion of diffusion distances between data samples as in Diffusion Maps, we work on the affinity matrix directly, using a self-induced smoothing kernel. Given data samples $\{x_i, \dots, x_n\}$ and a symmetric affinity function $\vartheta(x_i, x_j) = \vartheta(x_j, x_i) \in [0, 1]$, we define the $n \times n$ weight matrix W as $W(i, j) = \vartheta(x_i, x_j)$. SSO diffuses the pair-wise affinities of W along the geometry of the manifold without having to construct it explicitly:

1. Create the diffusion matrix $P = D^{-1}W$, where D is a diagonal matrix with $D_{ii} = \sum_{j=1}^n W(i, j)$.
2. Self diffusion: $W^{(p)} = WP^p$.

In Step 1, the smoothing kernel P that governs the diffusion process in Step 2 is induced from the input similarity matrix. $W^{(p)}$ is not a proper Gram matrix (since $\vartheta(x_i, x_j)$ is not an inner product), so it is in general neither symmetric nor positive semi-definite (PSD), which is not a problem in this application: we simply take the k minimal non-diagonal values of each row as the k nearest neighbors for classification. The only parameter in the algorithm is the step p , which determines the scale at which the data are analyzed. The output $W^{(p)}$ is an updated weight matrix that represents similarity more faithfully than W (as experimentally shown below) and that can be used directly in classification.

3.2 Similarity/Divergence Measures

The affinities $W(i, j)$ can be built from a similarity or divergence function, γ , using a suitable transform. If several $\{\gamma_m\}$, $m = 1, \dots, M$, are available, W can be a linear combination: $\bar{W} = w_m \sum_{m=1}^M W_m^{(p)}(\gamma_m)$, with $\sum_{m=1}^M w_m = 1$. We use two types of measures in this study: overlap- and registration-based. The first type is based on the Dice overlap of the structures of interest, giving a rough estimate of how similar two binary masks are. First, the centroids of the two instances of the structure to compare (s) are aligned, yielding $Y_{s,i}$ and $Y_{s,j}$. Then, the Dice overlap is computed as: $O(Y_{s,i}, Y_{s,j}) = \frac{2|Y_{s,i} \cap Y_{s,j}|}{|Y_{s,i}| + |Y_{s,j}|} \in [0, 1]$. One minus the overlap would be a valid affinity. However, in order to enhance



Fig. 1. (a) Slice of a sample volume, cropped around the hippocampus. The automatic hippocampal segmentation is outlined in black. The deformation field provided by the registration towards the slice in (b) is superimposed in blue. Summing the curvature (c) and diffusion (d) within the segmentation provides the divergences γ_{curv} and γ_{diff} .

the differences between good and bad matches, we linearly map the interval $[O_{s,min}, 1]$ to $[0, 1]$:

$$W_{over,s}(i, j) = 1 - \frac{O(Y_{s,i}, Y_{s,j}) - O_{s,min}}{1 - O_{s,min}} = \frac{1 - O(Y_{s,i}, Y_{s,j})}{1 - O_{s,min}}$$

A divergence function complementing the Dice coefficient should consider non-linear deformations. Here we use a diffeomorphic registration algorithm [18] to estimate the degree of warping that is required to deform a shape into another. To compare brains i and j , we first register j to i . Then, for structure of interest s , we compute the irregularity of the obtained deformation field $\mathbf{u}_{ji}(\mathbf{r})$ within the mask $\Omega_{s,i}$ corresponding to s in i (\mathbf{r} is the location vector). We use the curvature and diffusion of $\mathbf{u}_{ji}(\mathbf{r})$ as measures of irregularity:

$$\gamma_{curv}[\mathbf{u}(\mathbf{r})] = \int_{\Omega_{s,i}} \sum_{\{x,y,z\}} [\Delta u_d(\mathbf{r})]^2 d\mathbf{r}, \quad \gamma_{diff}[\mathbf{u}(\mathbf{r})] = \int_{\Omega_{s,i}} \sum_{\{x,y,z\}} \|\nabla_d(\mathbf{r})\|^2 d\mathbf{r}$$

where the index d loops along the three spatial dimensions. The deformation field for a sample case is shown in Fig. 1a. The integrands γ_{curv} and γ_{diff} are displayed in Fig. 1c and 1d. Finally, the corresponding weight matrices can be computed using a Gaussian function as follows:

$$W_{[\cdot]}(i, j) = \exp\left(-\left(\gamma_{[\cdot]}[\mathbf{u}_{ij}(\mathbf{r})] + \gamma_{[\cdot]}[\mathbf{u}_{ji}(\mathbf{r})]\right)^2 / (\text{var}(\gamma_{[\cdot]}))\right)$$

where $[\cdot]$ refers to curvature or diffusion, and $\text{var}(\gamma_{[\cdot]})$ is the variance of the divergence γ across the dataset. The weights $W_{[\cdot]}$ are explicitly symmetrized.

3.3 Feature Selection

Assuming that the global weight matrix \bar{W} is a linear combination of matrices based on single features (divergences or similarities), the question is which combination of weights $\mathbf{w} = \{w_m\}$ to use. Specifically, we seek to maximize the leave-one-out (LOO) classification rate $\Psi(\mathbf{w})$ under the constraints: $0 \leq \mathbf{w} \leq 1$ and $\mathbb{1}^t \mathbf{w} = 1$. This problem is difficult to solve because Ψ is neither smooth nor

convex, and has multiple local maxima. Instead, we further constrain the problem by assuming that only $M' \leq M$ weight matrices are used with equal weights $w_m = M'^{-1}, \forall m$. Then the problem becomes analogous to that of feature selection in machine learning. This is still a hard combinatorial problem, but good approximate solutions can be achieved using a proper selection strategy. Here we use “plus 2 - take away 1” [19]: from an initial empty set, features are greedily added / removed one at the time following the pattern $+, +, -, +, +, -, \dots$. The final set of features is the one that maximizes $\Psi(\mathbf{w})$ throughout the process.

4 Experiments and Results

4.1 Experimental Setup

The feature selection was cross-validated (10 folds) to obtain an unbiased estimate of the performance; otherwise features are selected upon the test data. For each fold, a set of features is selected with LOO on the training data. For each candidate set, the scale of the diffusion p is tuned individually using exhaustive search. The selected features and p are used to classify the test data in the fold. The number of neighbors was kept constant ($k = 10$) to limit the computational load of training. Ties are broken by examining subsequent neighbors. Rather than using all the 56 segmented structures in the selection process, only the (left and right) caudate nucleus, hippocampus and putamen are considered (18 features in total). These structures are well-known to be related to AD, and using this reduced set decreases the risk of overfitting.

In testing, an augmented W is created by adding to the original a new row and column for each test sample. We assume that all the test data are simultaneously available, which enables semi-supervised learning: during the diffusion process, the unlabeled test data can increase the performance of the system by making the structure of data easier to follow (only the labeled training data are considered during the k NN classification).

For the sake of comparison, analogous experiments were run using Diffusion Maps and the metric learning approach from [3], which attempts to find the positive definite matrix A that parameterizes the Mahalanobis distance best separating the training data into the different classes. Cross validation was again performed with 10 folds using the training data to select features (same selection strategy) and tune parameters: the matrix A for metric learning and the step t for Diffusion Maps (i.e. the scale of the diffusion). As for SSO, the number of nearest neighbors was fixed throughout the experiments ($k = 10$).

4.2 Results

The impact of feature selection on the performance is illustrated in Fig. 2a. The three most frequently selected features were: 1) diffusion - left hippocampus; 2) curvature - left putamen; and 3) overlap - left caudate. It is not surprising that the top feature is related to the hippocampus, which is known to be strongly connected with AD [5]. The curve in Fig. 2b shows the impact of the diffusion

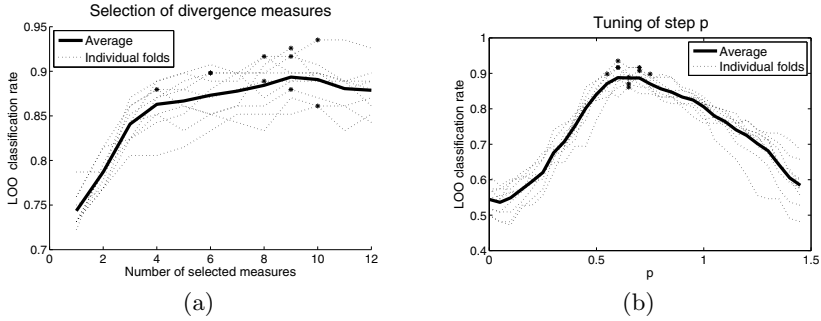


Fig. 2. (a) Best LOO classification rate against number of selected features. (b) For the feature subset chosen for each fold: dependence of the classification rate on the diffusion step p . The point $p = 0$ corresponds to the classification rate with no diffusion. In both graphs, the chosen operating points for each fold are marked with asterisks.

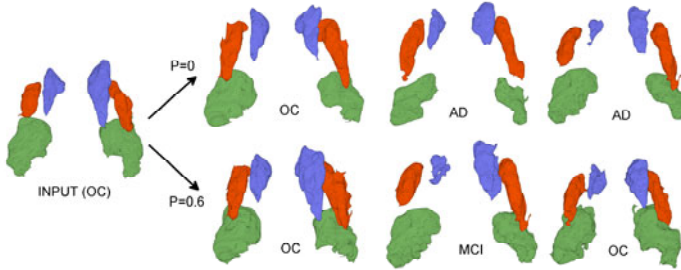


Fig. 3. Three-dimensional rendering of the structures of interest of an OC sample and its nearest neighbors, before ($p = 0$) and after diffusion ($p = 0.6$). The hippocampi are rendered in green, the putamens in red and the caudate nuclei in blue. The diffusion bridges the gaps with other OCs, moving the AD and MCI cases farther away.

on the classification. At first, increasing the scale of the diffusion p has a positive influence on the accuracy, which is boosted from $\sim 55\%$ at $p = 0$ (no diffusion) to $\sim 90\%$ at $p \approx 0.6$. This is illustrated with a sample subject and its nearest neighbors before and after diffusion in Fig. 3. When p becomes too large, data samples start to come too close to one another and the accuracy begins to decrease. Fortunately, the location of the peak is quite stable and the method generalizes well, as shown by the cross validation experiment below.

Tables 2a through 2c display the confusion matrices for metric learning, Diffusion Maps and our approach, respectively. Metric learning performs poorly because the structure of the data is too complex to discriminate the classes using a global linear transform. Diffusion Maps provides decent results: 78% accuracy with no mistakes between OC and AD. Our SSO-based approach makes no OC-AD mistakes either, but preserves the structure of the input similarity better than Diffusion Maps, increasing the accuracy to 89%. There is no

noticeable drop in accuracy from the training data (Fig. 2) because cross-validation (LOO) was already used within the feature selection process.

Even though the results reported in Table 1 were achieved on other datasets, it is illustrative to compare them to ours. *Chupin et al.*'s study, the only one considering the three-class problem, reports considerably lower accuracy than this work. To compare our results with the methods which classify OC vs. AD, we assume that only OC / AD are fed to the classifier and that the samples classified as MCI are relabeled to either OC or AD. Another option would be to remove the MCI cases from the training data, but that would have a negative impact on the results (the diffusion would be guided by less data). Our approach provides 96.25% or 97.5% accuracy (depending on the relabeling criterion), slightly higher than the best reported results in the literature (*Klöppel et al.*, 94%). In order to compare our approach with methods that discriminate OC from MCI, we assume that only OC and MCI cases are fed to the classifier, and the cases for which the estimated class is AD are relabeled as MCI. In that case, the accuracy is 91.25%, comparable to *Davatzikos et al.* and *Desikan et al.* (90%).

Table 2. Confusion matrices: (a) metric learning, (b) Diffusion Maps, (c) the proposed method. The global accuracies are 50%, 78% and 89%. GT stands for "ground truth".

(a)				(b)				(c)			
GT\Method	OC	MCI	AD	GT\Method	OC	MCI	AD	GT\Method	OC	MCI	AD
OC	20	11	9	OC	29	11	0	OC	38	2	0
MCI	14	18	8	MCI	5	29	6	MCI	5	32	3
AD	2	16	22	AD	0	4	36	AD	0	3	37

5 Discussion and Future Work

A nearest neighbor classifier based on registration and overlap features and enhanced by a self-smoothing operator has been presented in this study. SSO propagates the similarity between data samples along the manifold in which the data lie. The updated affinity measure can be used in a nearest neighbor framework to classify brains as AD, MCI or OC, achieving state-of-the-art results. The main disadvantage of the method is that, when a new case is presented to the system, computing the corresponding new row in the affinity matrix requires nonrigid registration to all the training cases, which is very time consuming (the SSO algorithm itself only takes a fraction of a second). Exploring its application to other disease patterns, testing features that are faster to compute and improving the design and combination of features remain as future work.

Acknowledgments. This work was funded by NSF career award IIS-0844566 and ONR grant N000140910099.

References

1. Alzheimer's Disease Neuroimaging Initiative, <http://adni.loni.ucla.edu>
2. Coifman, R., Lafon, S.: Diffusion maps. *Applied and Computational Harmonic Analysis* 21(1), 5–30 (2006)
3. Davis, J., Kulis, B., Jain, P., Sra, S., Dhillon, I.: Information-theoretic metric learning. In: *Proceedings of ICML*, pp. 209–216 (2007)
4. Karas, G., et al.: A comprehensive study of gray matter loss in patients with Alzheimer's disease using optimized voxel-based morphometry. *Neuroimage* 18(4), 895–907 (2003)
5. Du, A., et al.: Magnetic resonance imaging of the entorhinal cortex and hippocampus in mild cognitive impairment and Alzheimer's disease. *Journal of Neurology, Neurosurgery & Psychiatry* 71(4), 441–447 (2001)
6. Chupin, M., et al.: Fully automatic hippocampus segmentation and classification in Alzheimer's disease and mild cognitive impairment applied on data from ADNI. *Hippocampus* 19(6), 579–587 (2009)
7. Vemuri, P., et al.: Alzheimer's disease diagnosis in individual subjects using structural MR images: validation studies. *Neuroimage* 39(3), 1186–1197 (2008)
8. Klöppel, S., et al.: Accuracy of dementia diagnosis—a direct comparison between radiologists and a computerized method. *Brain* 131, 2969–2974 (2008)
9. Zhang, D., Wang, Y., Zhou, L., Yuan, H., Shen, D.: Multimodal classification of Alzheimer's disease and mild cognitive impairment. *NeuroImage* 55(3), 856–867 (2011)
10. Davatzikos, C., et al.: Detection of prodromal Alzheimer's disease via pattern classification of MRI. *Neurobiology of Aging* 29(4), 514–523 (2008)
11. Desikan, R., et al.: Automated MRI measures identify individuals with mild cognitive impairment and Alzheimer's disease. *Brain* 132(8), 2048 (2009)
12. Xing, I., Ng, A., Jordan, M., Russell, S.: Distance metric learning with application to clustering with side-information. *Advances in Neural Information Processing Systems*, 505–512 (2002)
13. Roweis, S., Saul, L.: Nonlinear dimensionality reduction by locally linear embedding. *Science* 290(5500), 2323
14. Gerber, S., Tasdizen, T., Fletcher, P., Joshi, S., Whitaker, R.: Manifold Modeling for Brain Population Analysis. *Medical Image Analysis* 14(5), 643–653 (2010)
15. Settles, B.: Active learning literature survey. *Computer Sciences Technical Report 1648*, University of Wisconsin–Madison (2009)
16. Smith, S.: Fast robust automated brain extraction. *Human Brain Mapping* 17(3), 143–155 (2002)
17. Tu, Z., et al.: Brain anatomical structure segmentation by hybrid discriminative/generative models. *IEEE Transactions on Medical Imaging* 27(4)
18. Avants, B., Epstein, C., Grossman, M., Gee, J.: Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis* 12(1), 26 (2008)
19. Stearns, S.: On selecting features for pattern classifiers. In: *Proc. of the 3rd Int. Joint Conf. on Pattern Recognition*, pp. 71–75

An Analysis of Scale and Rotation Invariance in the Bag-of-Features Method for Histopathological Image Classification

S. Hussain Raza¹, R. Mitchell Parry³, Richard A. Moffitt³, Andrew N. Young²,
and May D. Wang^{1,3,*}

¹ School of Electrical and Computer Engineering,
Georgia Institute of Technology, Atlanta, GA

² Pathology and Laboratory Medicine, Emory University, Atlanta, GA

³ The Wallace H. Coulter Department of Biomedical Engineering, Georgia Institute
of Technology and Emory University, Atlanta, GA
maywang@bme.gatech.edu

Abstract. The bag-of-features method has emerged as a useful and flexible tool that can capture medically relevant image characteristics. In this paper, we study the effect of scale and rotation invariance in the bag-of-features framework for Renal Cell Carcinoma subtype classification. We estimated the performance of different features by linear support vector machine over 10 iterations of 3-fold cross validation. For a very heterogeneous dataset labeled by an expert pathologist, we achieve a classification accuracy of 88% with four subtypes. Our study shows that rotation invariance is more important than scale invariance but combining both properties gives better classification performance.

Keywords: Bag-of-features Method, Texton-based Approach, Image Classification, Computer Aided Diagnosis.

1 Introduction

Renal Cell Carcinoma (RCC) accounts for 90-95% of adult malignancies arising from the kidney [1]. The American Cancer Society reported 58,240 new cases and 13,040 deaths in 2010 [1]. RCC occurs in four major subtypes: (i) Clear Cell (CC), (ii) Chromophobe (CH), (iii) Oncocytoma (ON), and (iv) Papillary (PA) [2]. Clinically, each subtype is treated differently. The task of subtype classification is performed by an expert pathologist under a microscope and suffers from subjectivity and observer variability [3]. Computerized histopathological image analysis aims at assisting a pathologist in the decision making process.

Recently, the bag-of-features approach has emerged as a useful tool for medical image classification [4-7]. The bag-of-features framework evolved from the bag-of-words model for text documents [8]. In the bag-of-words model, a dictionary is built from all the text documents and then each document is represented

* Corresponding author.

by the frequency of words in that document. The bag-of-features approach applies a similar methodology to image analysis. Images are divided into a collection of small patches, each of which described by a feature vector that encodes texture/content. Feature vectors are combined into a codebook that represents the characteristic patches in a collection of images. Typically, scale and rotation invariant features or raw pixel intensities are used [4-7]. Depending on the application, scenarios may exist where one or both could help or hurt performance. For example, cancer grading based on nucleus size may suffer from scale invariant features. However, there is no study exploring the impact of scale or rotation invariance for histopathological image analysis. In this paper, we perform an analysis of features with combinations of scale and rotation invariance in the bag-of-features framework. We have focused on the scale invariant feature transform (SIFT) to perform the analysis but other features such as speeded-up robust features (SURF) can also be used [9,10].

In histopathological image analysis, medically relevant morphologies can appear anywhere in the image and the spatial arrangement may not be important for decision making. Therefore, we do not consider spatial information preserving methods such as spatial pyramids [11]. In this paper, we evaluate the impact of scale and rotation invariance by studying the following types of features: (i) scale invariant features, (ii) rotation invariant features, (iii) features with both scale and rotation invariance, (iv) features with neither scale nor rotation invariance, and (v) raw pixel intensity based features. We evaluate their effect on histopathological image classification of RCC subtypes. This paper is organized as follows: section 2 provides background; section 3 provides methodology; section 4 shows the results; and section 5 concludes the analysis.

2 Background

A bag-of-features represents each image as a collection of features or patches. The relative abundance of each feature or patch distinguishes different types of images. These features or patches can be represented as invariant to scale and orientation. In this section, we provide the background of how this scale and rotation invariance is achieved. To achieve scale invariance, keypoints are selected by difference of Gaussian scale space (DoGSS) filtering [9]. The Gaussian scale space (GSS) of an image $I(x, y)$ is constructed by convolving that image with a Gaussian filter $G(x, y, \sigma)$ of different scales:

$$L(x, y, \sigma_i) = G(x, y, \sigma_i) * I(x, y),$$

where x and y are pixel coordinates, σ_i is the standard deviation of the Gaussian filter for scale i . We use previously reported values of $\sigma_i = k^i \sigma_0$, $\sigma_0 = 1.6$, and $k = \sqrt{2}$ [9]. Then DoGSS is computed by subtracting two consecutive images in the GSS:

$$D_i(x, y, \sigma_i) = L(x, y, \sigma_{i+1}) - L(x, y, \sigma_i) \quad (1)$$

Keypoints are detected by finding the extremas in DoGSS by comparing each pixel at D_i with its 3×3 neighborhood at scales D_i , D_{i-1} and D_{i+1} (i.e., 26

comparison in total). These keypoints appear in a single scale and additional scales can be achieved after downsampling and repeating [9]. A patch around each keypoint is encoded using SIFT to get a feature vector that represents the texture. To make the patch rotation invariant, the pixel coordinates x, y in the patch are rotated to align with the maximal gradient direction θ of that patch:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}, \quad (2)$$

where the pixel coordinates x, y are expressed with respect to the center of that patch. Another approach is to select the keypoints by dense sampling and encode the patch with raw pixel intensities to compute the features [4, 5]. After the features have been extracted from the images, codebook construction, image representation, and classification follows, as explained earlier (Figure 1).

3 Methods

The tissue samples are resected from renal tumors by total nephrectomy. Tissue samples of 3-millimeter thickness are obtained and fixed overnight in 10% neutral buffered formalin. Samples are then embedded in paraffin and microscopic sections of 5 micrometer thickness are prepared by a microtome and stained with hematoxylin and eosin. Photomicrographs of renal tumors are captured with 200x total magnification at 1200×1600 pixels per image. A total of 106

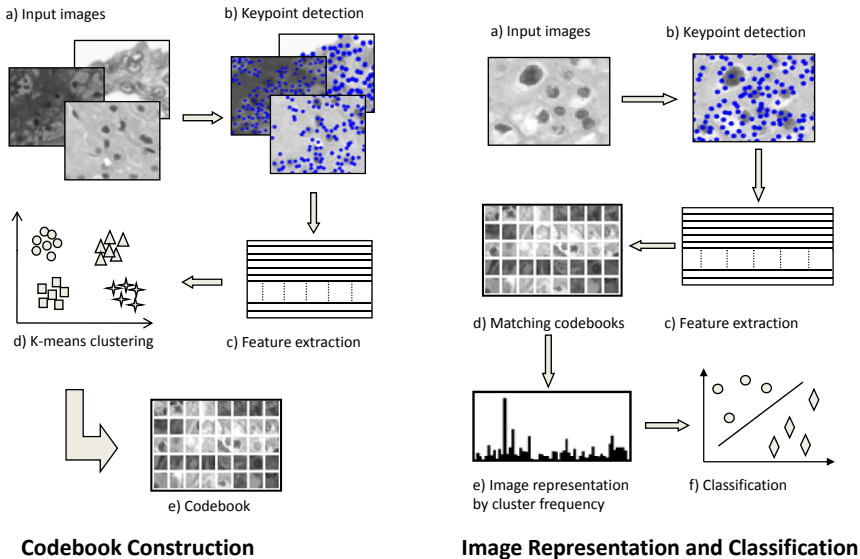


Fig. 1. Bag-of-features framework: (left) first, a codebook is constructed using k-means clustering over features extracted from training images only, (right) then, the codebook is used to generate a vector representation of each image using frequency of each cluster

images are captured, 32 of Chromophobe, 29 of Clear Cell, 28 of Papillary, and 17 of Oncocytoma. The images were labeled by an expert pathologist.

Figure 1 provides an overview of the bag-of-features approach. First, images are converted to gray scale and keypoints are selected to extract features. Features extracted from all the training images are used for constructing a codebook using k-means clustering (left of Figure 1). Next, a feature vector is constructed for each image by matching all the features from that image with the codebook (right of Figure 1). This feature vector is called a bag-of-features [8]. It represents the distribution of each cluster from the codebook in that image and does not account for their spatial relationship. A support vector machine (SVM) is trained with feature vectors from the training images and is used to classify the test images. Only the training set is used for learning the codebook and training the SVM.

To perform the analysis of scale and rotation invariance of features used in the bag-of-features framework for histopathological image classification, we adopted the following four strategies for keypoint selection and feature extraction based on the SIFT methodology. In the first strategy, we selected keypoints using DoGSS (Eq. 1) and computed SIFT descriptors invariant to orientation (Eq. 2) giving the standard SIFT features [9]. For DoGSS, we included the keypoints detected from scales D_{-1} to D_2 in our analysis. In each image about 6000 to 8000 keypoints were detected, 38% of total keypoints were detected at scale D_{-1} , 37% at scale D_0 , 13% at scale D_1 and, 12% were detected at scale D_2 . In the second strategy, we again included the keypoints from DOGSS but SIFT descriptors are computed by choosing a fixed orientation $\theta = 0$ in Eq. 2, resulting in only scale invariant features. In the third strategy, we computed the rotation invariant descriptors for a variety of fixed scales D_{-1} to D_2 . In the last strategy, we used the keypoints from Difference of Gaussian (DoG) at scale D_0 and computed SIFT descriptors using $\theta = 0$, giving features with fixed scale and orientation. Furthermore, we densely sampled 7000 keypoints from each image and computed rotation invariant features (for scales D_{-1} to D_2), features with fixed scale D_0 and orientation $\theta = 0$, and raw pixel intensity based features. Raw pixel intensity based features are computed over an area of 9×9 around the keypoint (i.e., fixed scale and orientation) [4, 5]. Table 1 summarizes keypoint detection and feature extraction for this paper.

Table 1. Summary of keypoint detection and feature extraction

Scale Inv.	Rotation Inv.	Keypoint Detection	Features
Yes	Yes	DoGSS	Scale & Rotation Invariance
Yes	No	DoGSS	Scale Invariance
No	Yes	DoG	Rotation Invariance
No	Yes	Dense sampling	Rotation Invariance
No	No	DoG	No Invariance
No	No	Dense sampling	No Invariance
No	No	Dense sampling	No Invariance (Raw Intensity)

We used a linear SVM with soft margin parameter $C = 1$ for classification [12]. Since SVM is a binary classifier, we adopted the “one vs. one with max voting” method to perform multiclass classification [12]. We performed 10 iterations of stratified 3-fold cross validation to estimate the performance.

4 Results and Discussion

Figure 2 illustrates the different types of features extracted with and without scale and rotation invariance. Figure 2a shows image patches with fixed scale D_0 and orientation $\theta = 0$. Because these patches have the same size (i.e., scale) and orientation, they don’t have the ability to match similar patches with different scale or orientation. Figure 2b shows patches with different scales but without rotation. Figure 2c shows image patches with rotation at a fixed scale D_0 . Figure 2d shows image patches with both scale and rotation invariance.

Figure 3 shows the performance comparison of different types of features in the bag-of-features framework for histopathological image classification of RCC subtypes. The plot shows that features with both scale and rotation invariance

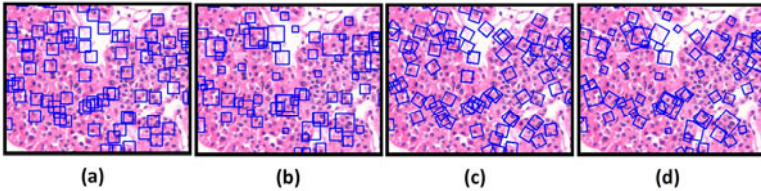


Fig. 2. The squares show the image patches selected for descriptor computation: (a) fixed scale D_0 and fixed orientation, (b) scale invariant, (c) fixed scale D_0 and rotation invariant, and (d) scale and rotation invariant.

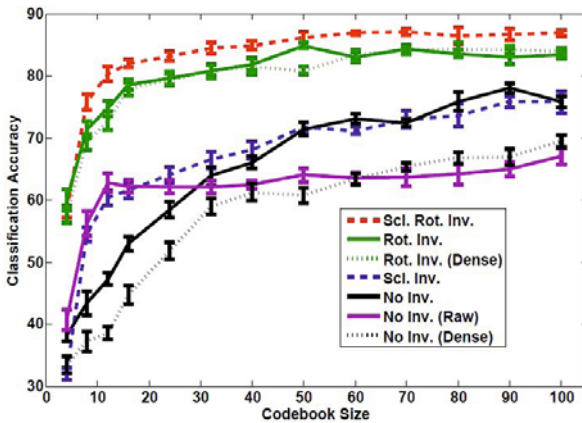


Fig. 3. Performance of different features in bag-of-features framework for histopathological image classification of RCC subtypes. Error bars show standard deviation of the means. Legend is in order with the curves in the plot at a codebook of size 45.

give better performance (top curve). A codebook of size 100 gives a classification accuracy of 88%. Features with rotation invariance also perform very well but features without rotation invariance perform poorly (lower four curves). For both dense sampling and fixed scale keypoint detection, we found that scale D_0 provided better top performance across codebook sizes and only plot D_0 for scale invariant performance in Figure 3. Specifically, rotation invariant features for dense sampling give a maximum classification accuracy of 84.1% at D_{-1} , 84.3% at D_0 , 82.9% at D_1 , and 80% at D_2 . For DoG keypoint detection, rotation invariant features give a maximum classification accuracy of 83.9%, 84.9%, 82.7%, and 81.2% at scales D_{-1} , D_0 , D_1 , and D_2 , respectively.

Figure 4 illustrates the difference of DoGSS and dense sampling keypoint selection strategies. The keypoints detected by DoGSS are dense in cellular regions and sparse in necrotic regions, whereas dense sampling selects keypoints that are dense in both cellular and necrotic regions. For RCC subtype classification, ignoring necrotic regions can improve classification accuracy [13]. We observe 5-10% improvement for features without scale and rotation invariance, partially confirming this finding. On the other hand, if the features are rotation invariant, we do not observe this effect. We speculate that smooth areas like necrotic regions have essentially random orientation. When the features are rotation invariant, the smooth necrotic regions get assigned to only a few clusters. However, when the features are not rotation invariant, the codebook must allocate a larger fraction of clusters to represent this randomness, thereby degrading performance. Table 2 gives the confusion matrices for each approach. It should be noted that incorporating the rotation invariance improves classification accuracy of each subtype as well as the overall classification accuracy. Furthermore, the confusion matrices for rotation invariant features are also very similar (Table 2d-e), i.e., both keypoint detection methods give similar performance for each subtype.

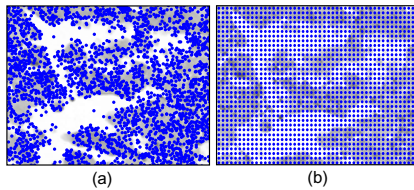


Fig. 4. Comparison of keypoint selection: (a) DoGSS, (b) Dense sampling

Another advantage of using scale and rotation invariant features is that the size of the codebook required to achieve good classification accuracy is smaller than codebooks developed by other features. To achieve classification accuracy over 80%, a codebook constructed over scale and rotation invariant features requires 12 types of patches, whereas rotation invariant features require 32 types of patches. Codebooks with just scale invariance require 200 types of patches to achieve 80% classification accuracy but a codebook of up to 300 patches without

Table 2. Average confusion matrices for six types of features over 10 iterations of 3-fold cross-validation for a codebook of size 40 clusters. Each row shows the true class labels and columns show predicted labels. Each row sums to the total number of images in that subtype.

	(a) No Inv.				(b) No Inv. (dense)				(c) Scale Inv.			
	CC	CH	ON	PA	CC	CH	ON	PA	CC	CH	ON	PA
CC	19.5	8.1	0.9	0.5	18.3	8.6	1.6	0.5	19.5	8.4	1.1	0.0
CH	10.0	18.2	2.9	0.9	8.4	18.9	3.0	1.7	8.6	19.2	3.0	1.2
ON	1.8	5.9	8.5	0.8	3.3	2.5	9.3	1.9	2.0	4.4	9.6	1.0
PA	1.5	1.2	1.5	23.8	3.6	3.2	2.8	18.4	1.3	1.0	1.9	23.8
	(d) Rot Inv.				(e) Rot. Inv. (dense)				(f) Scale & Rot. Inv.			
	CC	CH	ON	PA	CC	CH	ON	PA	CC	CH	ON	PA
CC	25.0	2.9	0.7	0.4	25.7	1.9	1.0	0.4	25.3	3.3	0.1	0.3
CH	5.5	25.4	1.1	0.0	4.3	25.6	1.7	0.4	3.5	25.9	2.1	0.5
ON	1.7	2.9	11.0	1.4	1.4	3.0	11.4	1.2	0.8	1.8	13.8	0.6
PA	0.7	0.1	1.8	25.4	1.1	0.7	2.5	23.7	1.4	0.3	1.3	25.0

scale or rotation invariance only achieves 75%. Features based on raw pixel intensities also performs poorly and a codebook of 300 patches gives a classification accuracy of about 70%. This suggest that given a large enough codebook, rotationally fixed features could eventually become equivalent to smaller rotationally invariant ones by encoding all of the possible different orientations an object can take. Figure 5 shows patches relevant to each RCC subtype identified by combining both scale and rotation invariance. Cyan squares show the cell membrane identified at different scales and orientations indicative of the clear cell subtype. Green squares show a complete round nucleus common in the oncocytoma subtype. Blue squares show a full nucleus with a halo around it indicative of the chromophobe subtype. Yellow squares show the streaks of finger-like structures characteristic of the papillary subtype.

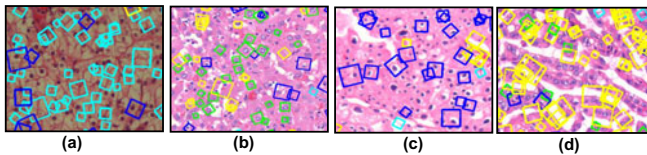


Fig. 5. Image patches matched with codebooks constructed over scale and rotation invariant features: (a) a cell membrane in cyan squares, (b) complete round nucleus in green squares, (c) complete nucleus with halo in blue squares, (d) and yellow squares showing finger-like structures

5 Conclusion

We conclude that rotation invariance is more important than scale invariance for histopathology image classification. Rotation invariant features computed at a

good choice of fixed scale perform nearly as well as scale and rotation invariant features. Therefore, covering the whole scale space may not be as important as choosing a preferred scale in histopathological image classification. Although rotation invariance combined with scale invariance performs slightly better, other applications may not benefit from this flexibility. For example, in the problem of cancer grading, nucleus size is important, and thus scale invariance may not be desirable. In the case of CT or MRI images, the pixel intensity is an important characteristic and a raw pixel intensity based feature could perform better than image gradient based features like SIFT. In the future, we would like to improve the codebooks by incorporating the class labels during the codebook generation and reduce the false positive matches by improving the distinctiveness of the features within the same codebook.

Acknowledgments. This research has been supported by grants from National Institutes of Health (Bioengineering Research Partnership R01CA108468, P20GM072069, Center for Cancer Nanotechnology Excellence U54CA119338), Georgia Cancer Coalition (Distinguished Cancer Scholar Award to Professor Wang), Hewlett Packard, and Microsoft Research.

References

1. National Cancer Institute - Kidney Cancer, <http://www.cancer.gov/cancertopics/types/kidney>
2. Atlas of Genetics and Cytogenetics in Oncology and Haematology, <http://www.AtlasGeneticsOncology.org>
3. Preti, M., Mezzetti, M., Robertson, C., Sideri, M.: Inter observer variation in histopathological diagnosis and grading of vulvar intraepithelial neoplasia: results of an European collaborative study. *BJOG: An International Journal of Obstetrics & Gynaecology* 107, 594–599 (2000)
4. Gangeh, M., Sørensen, L., Shaker, S., Kamel, M., de Bruijne, M., Loog, M.: A Texton-Based Approach for the Classification of Lung Parenchyma in CT Images. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) *MICCAI 2010*. LNCS, vol. 6363, pp. 595–602. Springer, Heidelberg (2010)
5. Avni, U., Greenspan, H., Sharon, M., Konen, E., Goldberger, J.: X-ray image categorization and retrieval using patch-based visualwords representation. In: *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pp. 350–353. IEEE press, Los Alamitos (2009)
6. Caicedo, J., Cruz, A., González, F.: Histopathology image classification using bag of features and kernel functions. In: Combi, C., Shaha, Y., Abu-Hanna, A. (eds.) *AIME 2009*. LNCS, vol. 5651, pp. 126–135. Springer, Heidelberg (2009)
7. Raza, S.H., Parry, R.M., Sharma, Y., Chaudry, Q., Moffitt, R.A., Young, A.N., Wang, M.D.: Automated Classification of Renal Cell Carcinoma Subtypes Using Bag-of-Features. In: *IEEE International Conference of Engineering in Medicine and Biology Society (EMBC)*, pp. 6749–6752. IEEE press, Los Alamitos (2010)
8. Csurka, G., Dance, C., Fan, L., Willamowski, J., Bray, C.: Visual categorization with bags of keypoints. In: *Workshop on Statistical Learning in Computer Vision (ECCV)*, vol. 01, p. 22. Citeseer (2004)

9. Lowe, D.: Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 74, 59–73 (2004)
10. Bay, H., Tuytelaars, T., Gool, L.V.: SURF: Speed Up Robust Features. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) *ECCV 2006*. LNCS, vol. 3951, pp. 404–417. Springer, Heidelberg (2006)
11. Liu, Y.Y., Chen, M., Ishikawa, H., Wollstein, G., Schuman, J., Rehg, J.: Automated Macular Pathology Diagnosis in Retinal OCT Images Using Multi-Scale Spatial Pyramid with Local Binary Patterns. In: Jiang, T., Navab, N., Plum, J.P.W., Viergever, M.A. (eds.) *MICCAI 2010*. LNCS, vol. 6361, pp. 1–9. Springer, Heidelberg (2010)
12. Manning, D.C., Raghavan, P., Schütze, H.: *Introduction to information retrieval*, pp. 328–330. Cambridge University Press, Cambridge (2008)
13. Chaudry, Q., Raza, S., Sharma, Y., Young, A., Wang, M.: Improving renal cell carcinoma classification by automatic region of interest selection. In: *8th IEEE International Conference on Bioinformatics and Bioengineering (BIBE)*, pp. 1–6. IEEE press, Los Alamitos (2008)

Robust Large Scale Prone-Supine Polyp Matching Using Local Features: A Metric Learning Approach

Meizhu Liu, Le Lu, Jinbo Bi, Vikas Raykar, Matthias Wolf,
and Marcos Salganicoff

Siemens Medical Solutions USA, Malvern, PA 19355

le-lu@siemens.com

University of Florida, Gainesville, FL 32611

mliu@cise.ufl.edu

Abstract. Computer aided detection (CAD) systems have emerged as noninvasive and effective tools, using 3D CT Colonography (CTC) for early detection of colonic polyps. In this paper, we propose a robust and automatic polyp prone-supine view matching method, to facilitate the regular CTC workflow where radiologists need to manually match the CAD findings in prone and supine CT scans for validation. Apart from previous colon registration approaches based on global geometric information [1-4], this paper presents a *feature selection* and *metric distance learning* approach to build a pairwise matching function (where true pairs of polyp detections have smaller distances than false pairs), learned using *local polyp classification features* [5-7]. Thus our process can seamlessly handle collapsed colon segments or other severe structural artifacts which often exist in CTC, since only local features are used, whereas other global geometry dependent methods may become invalid for collapsed segmentation cases. Our automatic approach is extensively evaluated using a large multi-site dataset of 195 patient cases in training and 223 cases for testing. No external examination on the correctness of colon segmentation topology [2] is needed. The results show that we achieve significantly superior matching accuracy than previous methods [1-4], on at least one order-of-magnitude larger CTC datasets.

1 Introduction

Colon cancer is the second leading cause of cancer death in western countries, but it is one of the most preventable of cancers because doctors can identify and remove its precursor known as a polyp. Besides the well established fiberoptic colonoscopy, 3D Computed Tomography Colonography (CTC), or virtual colonoscopy has emerged as a powerful screening tool for polyp detection and diagnosis. The research field of computer aided detection (CAD) of colonic polyps in CTC is highly exploited [5-7]. To enhance polyp findings in collapsed or fluid-tagged colon segments, and better distinguish polyps from pseudo polyps (e.g. tagged stools), the current CTC clinical practice is to obtain two scans of a

patient in prone and supine positions, respectively. However colon can move and deform significantly differently between the prone and supine scans, which makes the manual registration of polyp findings or colon segments difficult, inaccurate and time-consuming.

In this paper, we present a novel computerized technique to achieve high performance polyp matching, by supervisedly optimizing a distance metric in the feature space of polyp classification where true pairs of polyp matches statistically have smaller distance than false pairs. A polyp instance may be represented by a variety of *local appearance features* for classification [5-7], including local geometric features, and morphological, shape/intensity and context features. Since the total union of these features may lead to redundancy, greater computational and spatial complexity, we first use *feature selection* method to choose the features that are most relevant to polyp matching (e.g., the feature difference variation is minimal between true polyp matches), but least redundant. After selecting a subset of task-specific features, from the polyp classification feature pool, we propose an efficient *metric learning* method to learn a *covariance-matrix boosted Mahalanobis distance* to measure the instance differences across views.

Extensive evaluation is executed using a representative, multi-site clinical database with 195 patient cases in training and 223 cases for testing, containing 106, 118 polyps respectively. We demonstrate superior performance results on polyp prone-supine view matching, compared with existing work mostly based on colon centerline/surface registration [1-4]. Note that previous polyp matching techniques are tested and reported on datasets which are at least one order-of-magnitude smaller than ours, as 20 [1] (with 20 polyps), 12 [2] (with 12 polyps), and 39 [3] (with 23 training and 16 testing polyps) patients. This is partially because the pair of completely distended prone-supine colon scans (from rectum to cecum) is a prerequisite. Preparing topologically correct colon segmentation cases often needs manual editing or interactions [2] that can be labor-intensive, for large 3D volumes. Tagging residues or artifacts can largely affects the colon surface quality which imposes problems for global surface registration method [1].

2 Materials and Methods

Our approach consists of the following two steps. We first select a subset of features from the whole CAD classification feature pool, which is polyp matching-informative, using Minimum Redundancy Maximum Relevance (MRMR) algorithm [8]. Next, we learn an effective polyp matching distance metric on selected features (i.e., Mahalanobis distance by a positive semidefinite matrix that weights channels of features differently), in an additive, boosting based optimization manner. Fig. 1 summarizes the process diagram.

2.1 Matching-Sensitive Feature Selection

Data & Features: We collected 195 CTC cases (or 390 volumes) with 106 polyps appearing in both views for training; and 223 cases containing 118 polyps

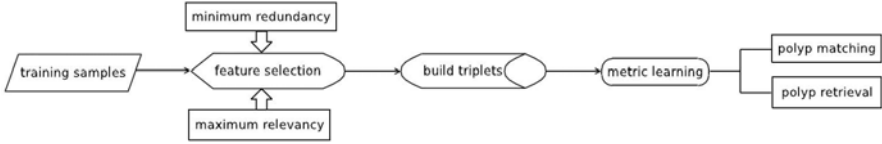


Fig. 1. The flow chart of using metric learning to do polyp matching and retrieval

with double appearance for testing, from 8 hospitals in US, Europe and Asia. Images are acquired using Siemens, GE and Philips scanners with fecal tagging preparation. Only actionable polyps with diameters $\geq 6\text{mm}$ are considered. Our CAD system builds a tree-structured probabilistic classifier using 96 (morphological, intensity, texture-based or geometrical) features $\mathcal{F} = \{f_i\}$ on 61257 candidates in training dataset. Thus we can first perform a thresholding process to rule out false positive (FP) candidates with low probability values (ρ of being polyp). After this, we have ~ 8 candidates per patient with true positive (TP) detection sensitivities at 94.2% and 92.9% for training and testing respectively. Note that our polyp matching approach is applicable on other CAD systems [5-7] which usually have a large variety of polyp descriptive features.

Let \mathbf{x}_i^1 be a true polyp instance in one view for a patient and let $\{\mathbf{x}_j^2\}$ be the set of corresponding instances in the other view. Note that the size of $\{\mathbf{x}_j^2\}$ can be larger than one since polyps can appear as two or more instances in each scan, especially for large polyps. This is called multiple instance problem. Here 1 or 2 indicates prone or supine view, without loss of generality. We define the positive (+) instance pairs of instances in two views rooted from the same unique true polyp, and other pairs as negative (-) (e.g., TP-TP pairs according to different polyps, TP-FP pairs, and FP-FP pairs). For each original feature f , a new “difference-of-feature” variable can be derived as $\Delta f = (f_i^1 - f_j^2)$, which is expected to be zero or a constant for positive pair population (i.e., tightly distributed in a statistical sense), or random for negatives.

Based on above motivation, we use feature selection algorithms to find a subset of $S \subset \mathcal{F}$ which are more informative on distinguishing true or false polyp pairs. For its numerical stability, accuracy and formulation simplicity, we choose Minimum Redundancy Maximum Relevance (MRMR) method though other feature selection methods are also applicable. For details, refer to [8]. As a result, we obtain 18 features out of 96 features, describing instance-level polyp’s shape morphology, segmented size, surface area, intensity profiles, classification score (“polypness”) and their joint statistics.

2.2 Matching by Metric Distance Learning

In this section, we propose a new metric learning method called “MatrixBoost” to match polyps in prone-supine views, using the 18 matching sensitive polyp features selected by MRMR. The basic idea is that a good distance metric can be learned to assign different weights on features, so that low distances are

given for pairs of instances to be matched and high distances for others, in the feature space. There are a number of ways to design distance metrics [9,11]. Metric learning can be derived from the optimal combination of weak learners to form a strong learner, based on training data. One type of weak learners is equivalence constrained, where equivalence constraints are provided for pairs $(\mathbf{x}_i, \mathbf{x}_j)$, each associated with a binary label of "similar" or "dissimilar" [9]. Another weak learner representation often used in information retrieval [11] is the proximity relationship over triplet set $\mathcal{T} = \{(i, j, k)\}$, meaning that \mathbf{x}_i is closer to \mathbf{x}_j than to \mathbf{x}_k . Here \mathbf{x}_i is the feature vector representation for the polyp instance i . The goal of metric learning is to learn a distance function d such that $d(\mathbf{x}_i, \mathbf{x}_j) < d(\mathbf{x}_i, \mathbf{x}_k)$, and

$$d(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i - \mathbf{x}_j)'M(\mathbf{x}_i - \mathbf{x}_j) \quad (1)$$

where $'$ is the vector/matrix transpose transformation and M is a positive semidefinite (PSD) matrix that leads to the Mahalanobis distance metric [9,10]. We follow the Mahalanobis distance metric formulation, but propose to construct the "covariance" matrix M by additively combining weak learners which are low rank PSD matrices. AdaBoost [12] method is utilized to learn the linear combination of low rank positive semidefinite (PSD) matrices, as a PSD matrix M preserving the proximity relationships among triplet set $\mathcal{T} = \{(i, j, k)\}$. The input to our metric boosting algorithm for training are triplets of instances, with inequality constraints on distances as defined above.

Build triplets. In training, we select all the instances with classifier score greater than $\rho \geq \beta_c$ to build triplets. The classifier score threshold β_c is chosen to make a practically feasible trade-off between detection sensitivity and FP rate (sensitivity is 0.946, and FP rate per patient is 7.59, pruning obvious negatives). The retained instances will form the triplets in the following way. For each true positive (polyp) instance \mathbf{x}_i in the prone view of a patient, we find all the positive instances $\{\mathbf{x}_j\}_{j=1}^n$ corresponding to the same polyp and all other instances (including positives corresponding to different polyps and negatives, or false positives) $\{\mathbf{x}_k\}_{k=1}^m$ in the supine view. Then (i, j, k) will form a triplet, requiring $d(\mathbf{x}_i, \mathbf{x}_j) < d(\mathbf{x}_i, \mathbf{x}_k)$. We repeat the same process on each true positive instance in the supine view to build more triplets, in a similar way. All the triplets form a triplet set \mathcal{T} and we obtained 8646 triplets in total, which will be used as inputs for our metric learning algorithm to optimize the PSD matrix M .

Learn a PSD matrix using MatrixBoost. Since a PSD matrix can be Eigen-decomposed as a combination of lower rank matrix, e.g., $M = \alpha_i U_i$, where $U_i = \mathbf{u}_i \mathbf{u}_i'$. The distance between two instances \mathbf{x}_i and \mathbf{x}_j is $d(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i - \mathbf{x}_j)'M(\mathbf{x}_i - \mathbf{x}_j)$. The algorithm is to learn a strong learner $H(\mathbf{x}, \mathbf{y}) = (\mathbf{x} - \mathbf{y})'M(\mathbf{x} - \mathbf{y})$, which is a combination of weak learners $h_t(\mathbf{x}, \mathbf{y}) = (\mathbf{x} - \mathbf{y})'U_t(\mathbf{x} - \mathbf{y})$, i.e. $H(\mathbf{x}, \mathbf{y}) = \sum_t \alpha_t h_t(\mathbf{x}, \mathbf{y})$, by minimizing the error rate of triplets violating the distance inequality as below.

$$\epsilon = \sum_{(i,j,k) \in \mathcal{T}} D((i, j, k)) \mathbf{1}_{(H(\mathbf{x}_i, \mathbf{x}_j) - H(\mathbf{x}_i, \mathbf{x}_k))}, \quad (2)$$

where D is a probability distribution over \mathcal{T} , and $\mathbf{1}$ is the Heaviside step function ($\mathbf{1}_{(a)} = 0$ if $a < 0$, and 1 otherwise). Our proposed MatrixBoost algorithm adapts the merits of the AdaBoost [12] and the decomposable nature of PSD distance matrix. In the algorithm, the weak model is $h_t(\mathbf{x}, \mathbf{y}) = (\mathbf{x} - \mathbf{y})'U_t(\mathbf{x} - \mathbf{y})$ where $U_t = \mathbf{u}_t\mathbf{u}_t'$, and the final hypothesis is $H(\mathbf{x}, \mathbf{y}) = (\mathbf{x} - \mathbf{y})'M(\mathbf{x} - \mathbf{y})$ where $M = \sum_t \alpha_t U_t$. Note that if M forms a metric that satisfies the triplet conditions, so does its multiplier. It can be proven that the training error of the final hypothesis H as defined in (2) is upper bounded by $\prod_{t=1}^T Z_t$, i.e.,

$$\begin{aligned} & \sum D((i, j, k)) \mathbf{1}_{(H(\mathbf{x}_i, \mathbf{x}_j) - H(\mathbf{x}_i, \mathbf{x}_k))} \\ & \leq \sum D((i, j, k)) \exp(H(\mathbf{x}_i, \mathbf{x}_j) - H(\mathbf{x}_i, \mathbf{x}_k)), \mathbf{1}_x \leq \exp(x) \\ & = \sum D_{\mathcal{T}+1}((i, j, k)) \prod_{t=1}^T Z_t = \prod_{t=1}^T Z_t. \end{aligned}$$

α_t and h_t will be chosen such that the error upper bound $\prod_{t=1}^T Z_t$ will be minimized. Let $h_t \in [0, 1]$, Z_t has the upper bound

$$Z_t \leq e^{\alpha_t} \frac{1-r}{2} + e^{-\alpha_t} \frac{1+r}{2}, \quad (3)$$

where

$$r = \sum_{(i,j,k) \in \mathcal{T}} D_t((i, j, k)) (h_t(\mathbf{x}_i, \mathbf{x}_k) - h_t(\mathbf{x}_i, \mathbf{x}_j)). \quad (4)$$

The right side of (3) can be minimized when $\alpha_t = \ln((1+r)/(1-r))/2$ which corresponds to $Z_t \leq \sqrt{1-r^2}$. Obviously, $Z_t \leq 1$ and if $r > 0$, we have $\alpha_t > 0$. Furthermore, the inequality implies that we can achieve smaller Z_t by minimizing its upper bound $\sqrt{1-r^2}$. Hence, a weak learner can be designed to maximize $|r|$ for a sensible model h_t .

$$\begin{aligned} & \max_{\substack{\mathbf{U}_t = \mathbf{u}_t\mathbf{u}_t' \\ \|\mathbf{u}_t\| = 1}} \left| \sum_{(i,j,k) \in \mathcal{T}} D_t((i, j, k)) (h_t(\mathbf{x}_i, \mathbf{x}_k) - h_t(\mathbf{x}_i, \mathbf{x}_j)) \right| \\ & \text{subject to } h_t(\mathbf{x}, \mathbf{y}) = (\mathbf{x} - \mathbf{y})' \mathbf{U}_t (\mathbf{x} - \mathbf{y}) \end{aligned} \quad (5)$$

Using simple matrix algebraic operations, Eq. (5) can be rewritten as

$$\left| \mathbf{u}_t' \left[\sum_{(i,j,k) \in \mathcal{T}} D_t((i, j, k)) ((\mathbf{x}_i - \mathbf{x}_k)(\mathbf{x}_i - \mathbf{x}_k)' - (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)') \right] \mathbf{u}_t \right| \quad (6)$$

The problem of maximizing the objective (6) subject to a normalization constraint $\|\mathbf{u}_t\| = 1$ has a closed-form solution: the optimal \mathbf{u}_t is the eigenvector corresponding to the eigenvalue λ , with the largest absolute value, of the matrix

$$\sum_{(i,j,k) \in \mathcal{T}} D_t((i, j, k)) ((\mathbf{x}_i - \mathbf{x}_k)(\mathbf{x}_i - \mathbf{x}_k)' - (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)'). \quad (7)$$

Let $a = \max\{\|\mathbf{x} - \mathbf{y}\| \mid \mathbf{x} \neq \mathbf{y}, \mathbf{x}, \mathbf{y} \in X\}$ which is a constant for a given set of data.

$$0 \leq h_t(\mathbf{x}, \mathbf{y}) = \frac{(\mathbf{x} - \mathbf{y})' \mathbf{u}_t \mathbf{u}_t' (\mathbf{x} - \mathbf{y})}{a^2} \leq \frac{\|\mathbf{x} - \mathbf{y}\|^2 \|\mathbf{u}_t\|^2}{a^2} \leq 1 \quad (8)$$

In testing, if $d(\mathbf{x}_i, \mathbf{x}_k) < \delta$ and δ is a distance threshold which can be statistically calibrated as shown later, we will claim \mathbf{x}_k is a match of \mathbf{x}_i and the confidence p_i for \mathbf{x}_i to be detected as positive (assuming only true positives have

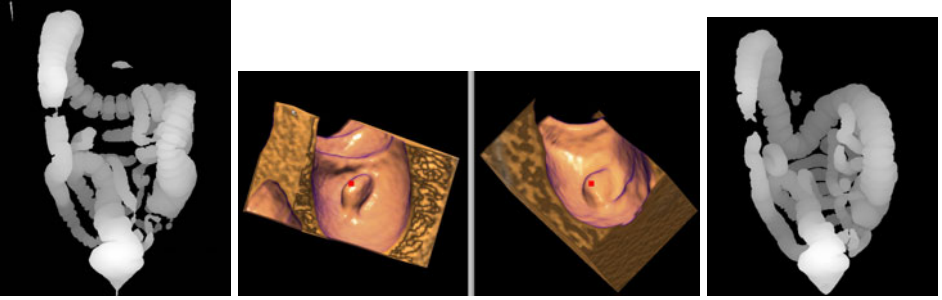


Fig. 2. An illustrative example of matched polyp pair, in a collapsed colon with background topological noises (e.g., small intestine) and large deformations in transverse and ascending colon sections

matches) is inverse to the minimum distance, i.e., $p_i = 1/d(\mathbf{x}_i, \mathbf{x}_k)$. An matched polyp pair example is shown in Fig. 2. The resulting weights or relative importances among selected features, reflected by the learned $M = \alpha_i U_i$, are well balanced among several cues of above local appearance. Even if some feature are forced to be removed, the matching performance degrades very gently.

3 Experimental Results

For each polyp \mathbf{x}_i in a given view (prone or supine), the goal of polyp retrieval is to find its corresponding counterpart in the other view. Using $d(\mathbf{x}_i, \mathbf{x}_j)$ in Eq. (1) with learned M , we sort its k nearest neighbors and check whether there is a true match within k to trigger a hit. The retrieval rate is defined as the number of polyps retrieved divided by the total number of query polyps. In case of multiple instances, any true instance appearing in the top k neighbors will count the polyp as retrieved at k . The evaluation of retrieval rate versus the number of nearest neighbors, i.e., k is demonstrated in Fig. 3 for both training and testing datasets. Fig. 3 shows the superior performance of metric learning methods (via fusing local polyp appearance features), compared with the centerline geodesic distance based retrieval, similar to [3, 4]. Not all polyps are retrievable because a small portion of polyps ($\leq 7\%$) only appeared in one view. For centerline based schemes, $> 40\%$ polyps are non-retrievable or can not be directly handled by [1-4], mainly due to collapsed colon segmentation in at least one prone or supine volume of 31% training, or 36% testing cases. [13] reports that $\sim 40\%$ volumes have collapsed colon segments in a clinical dataset. Note that, by normalizing against the polyp retrieval upper bounds (55 \sim 59% for geometric and 93 \sim 94% for metric learning) respectively, in Fig. 3, i.e., assuming all polyps are matchable, our *local features + metric distance learning* approach still clearly has more appealing performances, as 85% versus 62% in training; 80% versus 57% in testing when $k = 1$.

We also evaluate our MatrixBoost algorithm against other metric learning methods (Mahalanobis [10], PSDBoost [11], ITML [14], BoostMetric [11] and COP [9]), using the same selected feature set by MRMR. MatrixBoost dominates

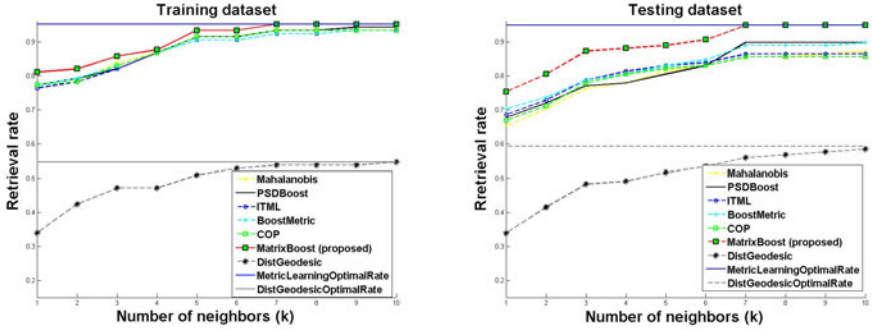


Fig. 3. Retrieval rates versus the number of nearest neighbors on the training (**Left**) and testing (**Right**) datasets

the retrieval rate at the full range of k , with a larger margin in testing dataset. For example, when $k = 2$, the testing retrieval rate of our method is 80.51%, while the best result of all other techniques is 73.73%. High polyp match/retrieval rates under smaller numbers of k , can greatly facilitate the workflow for radiologists to effectively and efficiently match the polyp findings in prone-supine CTC views. Moreover, MatrixBoost permits faster convergence to the upper bounds of polyp retrieval rate at $k = 7$ in both training and testing. Lastly, the polyp retrieval performance can be presented using Precision-Recall curves that show the balance of retrieval accuracy versus recall, in Fig. 4.

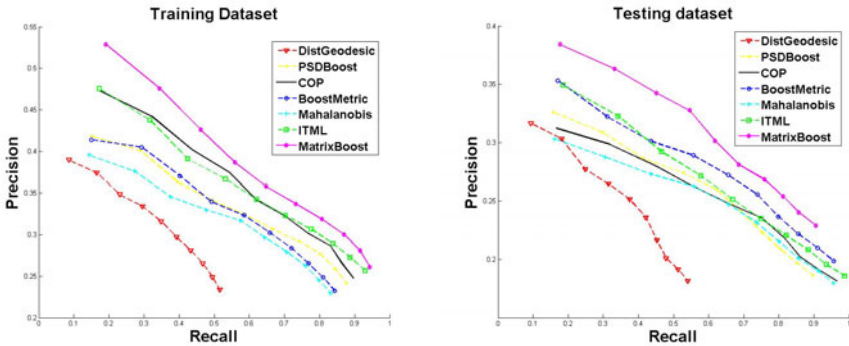


Fig. 4. The comparison of polyp retrieval Precision-Recall curves, using different metric learning methods and centerline Geodesic distance approach, on training (**Left**) and testing (**Right**) datasets. MatrixBoost method shows superior performance than others.

4 Discussion

We proposed an effective and high performance polyp prone-supine view matching method, based on local polyp classification feature learning (via feature se-

lection and metric learning). Our approach is evaluated on at least one order-of-magnitude larger, multiple hospitals dataset than previous work [14]. It can automatically and robustly handle highly varying colon segmentations from hundreds of patient cases, without any manual editing or preprocessing overhead. In summary, our method greatly advances the state-of-the-arts for polyp matching, and makes it more technically feasible for clinical practice.

References

1. Lai, Z., Hu, J., Liu, C., Taimour, V., Pai, D., Zhu, J., Xu, J., Hua, J.: Intra-patient Supine-Prone Colon Registration in CT Colonography Using Shape Spectrum. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6361, pp. 332–339. Springer, Heidelberg (2010)
2. Roth, H., McClelland, J., Modat, M., Boone, D., Hu, M., Ourselin, S., Slabaugh, G., Halligan, S., Hawkes, D.: Establishing Spatial Correspondence between the Inner Colon Surfaces from Prone and Supine CT Colonography. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6363, pp. 497–504. Springer, Heidelberg (2010)
3. Wang, S., Yao, J., Liu, J., Petrick, N., Periaswamy, S., Summers, R.: Registration of Prone and Supine CT Colonography Scans Using Correlation Optimized Warping and Canonical Correlation Analysis. *Medical Physics* 36, 5595–5603 (2009)
4. Li, P., Napel, S., Acar, B., Paik, D., Jeffery, R., Beaulieu, C.: Registration of Central Paths and Colonic Polyps Between Supine and Prone Scans in Computed Tomography Colonography: Pilot Study. *Medical Physics* 31, 2912–2923 (2004)
5. Mang, T., Bogoni, L., Salganicoff, M., Wolf, M.: CT Colonography: Retrospective Evaluation of the Performance of Computer-aided Detection of Colonic Polyps in Tagged and Untagged Preparation. In: European Congress of Radiology (2010)
6. Slabaugh, G., Yang, X., Ye, X., Boyes, R., Beddoe, G.: A Robust and Fast System for CTC Computer-Aided Detection of Colorectal Lesions. *Algorithms* 3(1) (2010)
7. van Ravesteijn, V., van Wijk, C., Vos, F., Truyen, R., Peters, J., Stoker, J., van Vliet, L.: Computer Aided Detection of Polyps in CT Colonography using Logistic Regression. *IEEE Trans. on Med. Imag.* 29, 120–131 (2010)
8. Peng, H., Long, F., Ding, C.: Feature Selection Based on Mutual Information: Criteria of Max-dependency, Max-relevance, and Min-redundancy. *IEEE TPAMI*, 1226–1238 (2005)
9. Xing, E., Ng, A., Jordan, M., Russell, S.: Distance Metric Learning, with Application to Clustering with Side-information. In: NIPS, vol. 15, pp. 505–512 (2002)
10. Bar-Hillel, A., Hertz, T., Shental, N., Weinshall, D.: Learning a Mahalanobis Metric from Equivalence Constraints. *J. Mach. Learn. Res.* 6, 937–965 (2005)
11. Shen, C., Kim, J., Wang, L., van den Hengel, A.: Positive Semidefinite Metric Learning with Boosting. In: NIPS, pp. 1651–1659 (2009)
12. Freund, Y., Schapire, R.: A Decision-Theoretic Generalization of on-Line Learning and an Application to Boosting. *Comput. Learn. Theory* (1995)
13. Lu, L., Wolf, M., Liang, J., Bi, J., Salganicoff, M.: A Two-Level Approach Towards Semantic Colon Segmentation: Removing Extra-Colonic Findings. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5762, pp. 1009–1016. Springer, Heidelberg (2009)
14. Davis, J., Kulis, B., Jain, P., Sra, S., Dhillon, I.S.: Information-Theoretic Metric Learning. In: ICML, pp. 209–216 (2007)

Targeted Optical Biopsies for Surveillance Endoscopies

Selen Atasoy^{1,2}, Diana Mateus¹, Alexander Meining³,
Guang-Zhong Yang², and Nassir Navab¹

¹ Chair for Computer Aided Medical Procedures (CAMP), TU Munich, Germany
`{atasoy,mateus,navab}@cs.tum.edu`

² Hamlyn Centre for Robotic Surgery, Imperial College London, United Kingdom
`{catasoy,gzy}@doc.ic.ac.uk`

³ Klinikum Rechts der Isar, TU Munich, Germany
`alexander.meining@lrz.tu-muenchen.de`

Abstract. Recent introduction of probe-based confocal laser endomicroscopy (pCLE) allowed for the acquisition of *in-vivo* optical biopsies during the endoscopic examination without removing any tissue sample. The non-invasive nature of the optical biopsies makes the re-targeting of previous biopsy sites in surveillance examinations difficult due to the absence of scars or surface landmarks. In this work, we introduce a new method for recognition of optical biopsy scenes of the diagnosis endoscopy during serial surveillance examinations. To this end, together with our clinical partners, we propose a new workflow involving two-run surveillance endoscopies to reduce the ill-posedness of the task. In the first run, the endoscope is guided from the mouth to the z-line (junction from the oesophagus to the stomach). Our method relies on clustering the frames of the diagnosis and the first run surveillance (S_1) endoscopy into several scenes and establishing cluster correspondences across these videos. During the second run surveillance (S_2), the scene recognition is performed in *real-time* and *in-vivo* based on the cluster correspondences. Detailed experimental results demonstrate the feasibility of the proposed approach with 89.75% recall and 80.91% precision on 3 patient datasets.

1 Introduction

Oesophageal adenocarcinoma (OAC) is one of the most rapidly increasing cancers in the Western world with a survival rate of less than 20%. The reason of this low survival rate in OAC is largely due to its late diagnosis. To alleviate this problem, patients diagnosed with a precursor of OAC are required to undergo regular surveillance endoscopies where biopsies are taken from suspicious tissue regions. The introduction of the new probe-based confocal laser endomicroscopy (pCLE) enabled real-time visualisation of cellular structures *in-vivo*. Despite their established advantages, these *optical biopsies* also introduce new challenges into the existing gastro-intestinal (GI) endoscopy workflow. Due to their non-invasive nature, re-targeting the same biopsy locations in subsequent surveillance examination becomes very challenging. Recently, several methods have been proposed

for addressing the re-localization problem within one intervention [2,11,13]. The application of such localization methods to a *new surveillance* GI endoscopy requires real-time recognition of the frames containing previously targeted biopsy sites. The major challenge of performing scene recognition between the diagnosis and surveillance endoscopies is the variation in visual appearances of the same scene as demonstrated in Fig. 1(a),(d). To address this challenge, we propose a *two-run surveillance endoscopy*. In the introduced workflow, prior to the actual surveillance endoscopy, a first-run surveillance ($\mathcal{S}1$) video is acquired in the same examination. This is a commonly performed process in bronchoscopy [9]. To the best of our knowledge, however, this process has not been applied in GI examinations. In this work, we introduce the two run surveillance schema for GI endoscopies, which allows us to provide an applicable solution for re-targeting the optical biopsy sites in surveillance examinations.

The proposed method first creates scene clusters from the diagnosis and $\mathcal{S}1$ endoscopies and then establishes correspondences between these two videos based on expert’s supervision. As the structure of the tissue between the $\mathcal{S}1$ and the actual examination performed in the second run surveillance ($\mathcal{S}2$) remains the same, the visual recognition of a scene becomes a solvable task. Once the query scenes, *i.e.* scenes of the diagnosis endoscopy which need to be recognized, are defined, recognition is achieved based on the guided correspondences.

To facilitate the proposed workflow, an endoscopic scene clustering method proposed in [4] is adapted. To this end, we create a manifold representation of the endoscopic videos by taking into account the visual similarities and the temporal relations within the video simultaneously. Scene clustering is performed in the low dimensional space using a mixture model method presented in [7]. The accuracy of the method is validated on 3 different patient datasets, where the patient underwent chemotherapy between the acquisitions.

2 Methods

2.1 Proposed Workflow

In this work, we firstly propose a *two-run surveillance endoscopy*. In the introduced schema, prior to the actual surveillance endoscopy, the endoscope is guided from the mouth to the z-line (junction from the oesophagus to the stomach) without acquiring any optical biopsies. The video of this $\mathcal{S}1$ endoscopy is clustered into different endoscopic scenes and used to acquire scene matching between the diagnosis and surveillance endoscopy. This additional step enables the recognition of the same location despite very large variation in the visual appearances of the scene in different examinations as illustrated in Fig. 1(a),(d).

Thus, the proposed workflow involves 3 endoscopic videos: diagnosis endoscopy (Fig. 1(a)), where the first optical biopsies have been acquired; $\mathcal{S}1$ (Fig. 1(d)) which is performed to provide matches between the endoscopic scene clusters; and the $\mathcal{S}2$ (Fig. 1(g)) where the surveillance examination is performed and the previous optical biopsy sites need to be recognized in real-time and *in-vivo*.

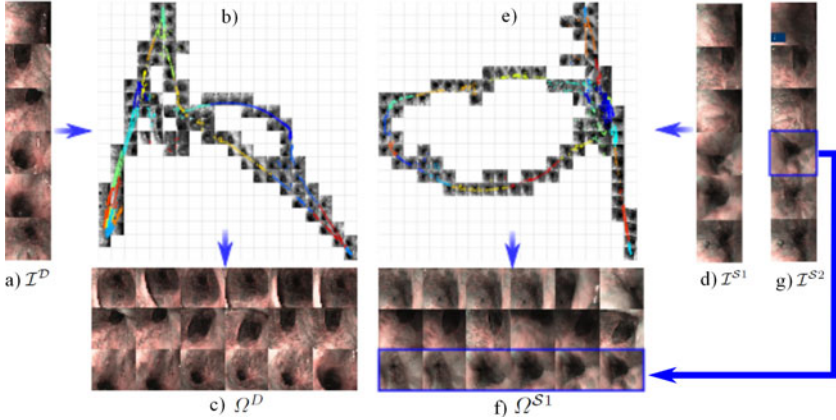


Fig. 1. Proposed workflow. a) Frames from the diagnosis endoscopy. b) 1. and 2. dimensions of the manifold of the diagnosis endoscopy created using vtLPP. Frames showing similar locations are clustered together, where clusters are illustrated with different colors. c) Example clusters of the diagnosis endoscopy, where rows correspond to different clusters. Note that frames of the same scene with different endoscope viewpoint are clustered together whereas different scenes are clustered separately. d) Corresponding scenes of a) in the $S1$. Rows in a) correspond to rows in d). e) 1. and 2. dimensions of $S1$ manifold and the computed clusters. f) Frames from the corresponding clusters of c) in the $S1$. The rows in c) correspond to rows in f). g) Example frames from the $S2$.

The proposed workflow consists of the following main steps:

1. Clustering of the diagnosis endoscopy into different scenes (Fig. 1(a)-(c)),
2. Acquisition of the $S1$ endoscopy (Fig. 1(d)),
3. Clustering of the $S1$ endoscopy into different scenes (Fig. 1(d)-(f)),
4. Selection of the query clusters in the diagnosis endoscopy and their correspondences in the $S1$ by the endoscopic expert,
5. Nearest neighbour matching and $S1$ cluster assignment to each frame of the $S2$ endoscopy in real-time (Fig. 1(g)),
6. Notification of the expert during the $S2$ endoscopy if a frame is assigned to one of the query clusters.

Given the frames of the diagnosis (Fig. 1(a)) and of the $S1$ (Fig. 1(d)) endoscopies, our method first computes a low dimensional manifold representation for each video by taking into account the visual similarities and the temporal relations between the frames. This allows for efficient clustering of the endoscopic scenes. Fig. 1(b) and (e) show the 1. and 2. dimensions of the manifolds computed from the diagnosis and $S1$ endoscopies respectively, where the clusters are illustrated by different colors. Clustering of the frames into different scenes is performed on this manifold representation using a mixture model and the expectation maximization method proposed in [7]. Fig. 1(c) shows example clusters from the diagnosis endoscopy where the corresponding clusters in the $S1$ are illustrated in Fig. 1(f). Note the severe change in the appearance of the

scenes between the two examinations. Based on the previously defined diagnosis endoscopy clusters and their correspondences in the $\mathcal{S}1$, the proposed workflow allows for *real-time* and *in-vivo* recognition of the query scenes during the $\mathcal{S}2$.

2.2 Data Representation

Clustering of endoscopic frames using the original image representation is not practical due to the high dimensionality of the data. In [4], the authors propose to recover the underlying non-linear manifold structure of an endoscopic video and to perform the clustering on this low dimensional space. In this work, we approximate the manifold underlying an endoscopic video using the locality preserving projections (LPP) method [10]. In contrast to [4], we define the relations between the frames by taking into account their visual similarities and temporal relations simultaneously and use a probabilistic clustering presented in [7].

LPP first defines an adjacency graph A that captures the pairwise relations $A(i, j)$ between the frames \mathcal{I}_i and \mathcal{I}_j , ($i, j \in \{1, \dots, n\}$, n being the number of data points), and then estimates a mapping to embed the graph into a low dimensional space. In order to simultaneously capture the visual and the temporal relations between the data points, we propose to define the adjacency graph as:

$$A(i, j) = \begin{cases} 1 & \text{if } i \in \mathcal{N}_j^{\text{sim}} \text{ or } i \in \mathcal{N}_j^{\text{temp}} \\ 0 & \text{otherwise} \end{cases}, \quad (1)$$

where $\mathcal{N}_j^{\text{sim}}$ is the k -NN of the j -th data point based on the visual similarities and $\mathcal{N}_j^{\text{temp}}$ states the k -NN based on the temporal order of the frames within the endoscopic video. In this work, we determine the visual similarities using the Euclidean distance and choose $k = 20$ considering the observed endoscope motion. Imposing the proposed temporal constraint assures that frames showing the same scene from different endoscope viewpoints are closely localized on the manifold, even in cases where visual similarities fail to capture their relations. On the other hand, using the visual similarities includes the neighborhood of similar but temporally distant frames, which is reflected in the closed loops on the manifold representations (Fig. 1(b),(e)).

Given the adjacency matrix A and the (vectorized) endoscopic frames $\mathcal{I} = [\mathcal{I}_1, \mathcal{I}_2, \dots, \mathcal{I}_n]$, we approximate the underlying manifold of the endoscopic data using the LPP method [10]. In LPP, first a function basis $w = [w_1, \dots, w_m]$ is computed based on locally linear approximations of the Laplace-Beltrami operator applied on the dataset by solving the following eigenvalue problem:

$$\mathcal{I}L\mathcal{I}^\top w = \lambda \mathcal{I}D\mathcal{I}^\top w, \quad (2)$$

where D is the diagonal degree matrix with $D(i, i) = \sum_j A(j, i)$ and $L = D - A$ is the graph Laplacian matrix [10]. Then the m dimensional representation $\nu = [\nu_1(i), \dots, \nu_m(i)]^\top$ of a frame \mathcal{I}_i is estimated by projecting it onto the estimated basis $\nu = w^\top \mathcal{I}_i$. Thus, this method provides an approximation for the Laplacian Eigenmaps (LE) method [5] while it also allows for projection of new data points

onto the manifold. Fig. III(b),(e) illustrate a 2D representation of two endoscopic videos. In the rest of the paper, we refer to our representation as visual and temporal LPP (vtLPP).

2.3 Endoscopic Scene Clustering

Once the low dimensional representations of endoscopic frames are computed, we use the finite mixture models (FMM) method proposed in [7] to compute the clusters. Using FMM, we estimate the probability $P[c(\nu(i)) = C_j]$ of each point $\nu(i)$ belonging to a mixture model (cluster) C_j and assign the cluster with the highest probability $c(\nu(i)) = \arg \max_{C_j} P[c(\nu(i)) = C_j]$. FMM [7] offers the advantage of automatically detecting the number of clusters. Additionally, FMM models clusters with anisotropic Gaussians, which overcomes the isotropic distribution assumption imposed in clustering algorithms such as K-means [8] and results in elongated clusters. Such clusters efficiently group frames showing the same scene with different viewpoints as shown in Fig. III(b),(c),(e) and (f).

2.4 Endoscopic Scene Recognition

After computing the clusters of the diagnosis endoscopy $\Omega^{\mathcal{D}} = \{C_1^{\mathcal{D}}, \dots, C_\alpha^{\mathcal{D}}\}$ and then the ones of the S1 endoscopy $\Omega^{\mathcal{S}1} = \{C_1^{\mathcal{S}1}, \dots, C_\beta^{\mathcal{S}1}\}$, both clusterings are provided to the endoscopic expert. The set of Q clusters, where an automatic recognition is needed, *i.e.* the query clusters $\{C_q^{\mathcal{D}}\}_{q=1}^Q \in \Omega^{\mathcal{D}}$, as well as their correspondences in the S1 endoscopy, $\{C_{\gamma(q)}^{\mathcal{S}1}\} \in \Omega^{\mathcal{S}1}$ (where γ denotes the correspondence relation) are selected by the endoscopic expert.

During the S2, first the image closest to a frame $\mathcal{I}_i^{\mathcal{S}2}$, that is $\mathcal{I}_j^{\mathcal{S}1} = NN(\mathcal{I}_i^{\mathcal{S}2})$, is found by a simple NN matching using Euclidean distances. Then each frame $\mathcal{I}_i^{\mathcal{S}2}$ is assigned the cluster of its NN $c^{\mathcal{S}1}(\mathcal{I}_i^{\mathcal{S}2}) = c^{\mathcal{S}1}(\mathcal{I}_j^{\mathcal{S}1})$ and, by transition, the corresponding diagnosis endoscopy cluster $c^{\mathcal{D}}(\mathcal{I}_i^{\mathcal{S}2}) = c^{\mathcal{D}}(\mathcal{I}_j^{\mathcal{S}1})$. If a frame is determined to belong to a query cluster $c^{\mathcal{D}}(\mathcal{I}_i^{\mathcal{S}2}) \in \{C_q^{\mathcal{D}}\}$, the expert is notified and all frames of the corresponding diagnosis endoscopy cluster $\{\mathcal{I}_k^{\mathcal{D}} | c^{\mathcal{D}}(\mathcal{I}_k^{\mathcal{D}})\}$ are retrieved. This proposed workflow thus allows for including the expert's supervision in defining the query scenes and their correspondences in the S1 without involving any training. This is an important property, since long training processes would not be feasible for routine clinical applications.

3 Experiments and Results

Experiments were performed on 3 narrow-band imaging (NBI) patient datasets acquired at 3 different examinations of the same patient. The patient underwent chemotherapy between the examinations, leading to significant changes in the appearance of the tissue as illustrated in Fig. III. Uninformative frames are labeled using the method in [4] and the remaining informative frames (1198, 1833 and 712 frames in 1., 2. and 3. datasets, respectively) are used for the experiments.

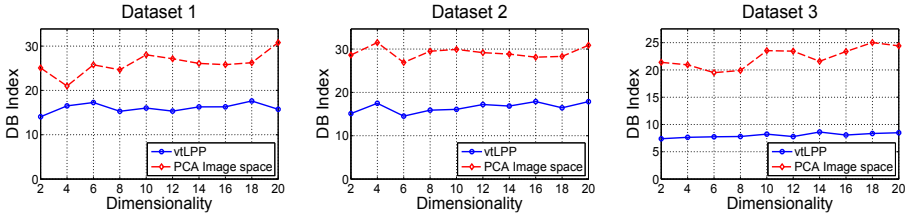


Fig. 2. Evaluation of scene clustering on the proposed representation as compared to the low dimensional image space representation

3.1 Evaluation of Scene Clustering

In order to assess the quality of the clustering, we evaluate the Davis-Bouldin (DB) index [6] which is a commonly used evaluation criteria for clustering algorithms. DB-index measures the relation of the between cluster distances (separability) and within cluster distances (compactness) and is independent of the number of clusters. Smaller DB-indices indicate more compact and separable clusters and are desired. We compare the DB-index of the clustering performed in our vtLPP representation to the one in the PCA representation of the data. Due to its numerical instability, the FMM algorithm [7] is not applicable to very high dimensional data, such as in the original image representation. Therefore, we apply a principal component analysis (PCA) and reduce the dimensionality of the dataset prior to clustering. Using the FMM clustering in [7], we observed that higher dimensional representations result in less number of clusters. Therefore, the evaluation of the DB-index is performed by varying the dimensionality from 2 to 20 for the two methods. Fig. 2 shows that for all number of dimensions and for all datasets, the proposed representation results in significantly smaller DB-indices indicating more compact and better separated clusters.

3.2 Evaluation of Scene Recognition

For quantitative analysis we perform 3 experiments. In each experiment, 40 frames from the surveillance endoscopic video are selected by regularly sampling the frames over time and are used as test frames simulating the S_2 endoscopic frames leading to a total recognition of 120 frames. Remaining parts of the surveillance video are defined to be the S_1 endoscopy. The results are compared to k -NN matching based on Euclidean distances performed between the S_2 and diagnosis endoscopy frames directly, where k is chosen to be equal to the number of frames retrieved by our method. We also performed the NN matching using the normalized cross correlation and did not observe a significant improvement in the recognition results. The true positives (tp) and false positives (fp) are determined by expert visual inspection of the retrieved frames. The false negatives (fn) of each method is defined relatively, as the number of frames that one method is able to correctly retrieve but not the other. Recall ($tp/(tp + fn)$) and precision ($tp/(tp + fp)$) values are evaluated for each test frame and mean and standard deviation achieved by both methods is presented in Fig. 3. Application

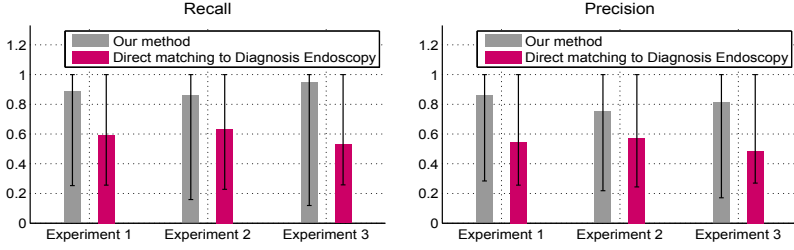


Fig. 3. Mean and standard deviation of recall and precision of the proposed method and of the direct application of the k -NN matching to the diagnosis endoscopy

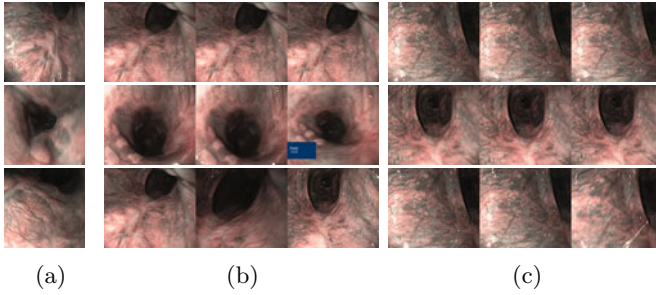


Fig. 4. a) Test frames used as $S2$ endoscopy. b) Recognized frames using our method. c) 3 NN in the diagnosis endoscopy. The rows show corresponding frames in a), b), c).

of the k -NN matching directly between the test frames and the diagnosis endoscopy results in only 58.54% mean recall and 53.58% mean precision. Our proposed method leads to a 89.75% recall and 80.91% precision in average using the same NN matching between the test frames and the $S1$ endoscopic frames and then applying the cluster correspondences. Examples of the correctly recognized frames using the proposed method in comparison to the direct application of k -NN matching between the $S2$ and diagnosis videos are demonstrated in Fig. 4. Due to the use of our vtLPP representation, the formed endoscopic clusters contain frames showing the same location from different viewpoints and from different parts of the video. This is also reflected in the high recall and precision values of the proposed method.

4 Conclusions

In this work, we present an endoscopic scene recognition method based on two run surveillance endoscopies and scene clustering. The key contributions of this work are two-fold. Technically, we have presented a scene clustering method for endoscopic videos by taking into account both visual similarities and temporal relations in a low dimensional space. Clinically, we have proposed a solution to the challenging problem of re-targeting the optical biopsy sites in *surveillance*

endoscopies. The introduced workflow allows us to create a link between the scenes of the diagnosis and surveillance examinations. This reformulation reduces the very challenging inter-examination re-targeting into the plausible problem of intra-examination frame recognition. The experiments on 3 different patient datasets demonstrate the feasibility of our method to recognize the optical biopsy scenes in surveillance endoscopies.

Acknowledgements. This research was supported by the Graduate School of Information Science in Health (GSISH) and the TUM Graduate School. The authors would like to thank Alessio Dore for valuable discussions.

References

1. Allain, B., Hu, M., Lovat, L., Cook, R., Ourselin, S., Hawkes, D.: Biopsy Site Re-localisation Based on the Computation of Epipolar Lines from Two Previous Endoscopic Images. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009, Part I. LNCS, vol. 5761, pp. 491–498. Springer, Heidelberg (2009)
2. Allain, B., Hu, M., Lovat, L., Cook, R., Vercauteren, T., Ourselin, S., Hawkes, D.: A System for Biopsy Site Re-targeting with Uncertainty in Gastroenterology and Oropharyngeal Examinations. In: Jiang, T., Navab, N., Pluim, J., Viergever, M.A. (eds.) MICCAI 2010, Part II. LNCS, vol. 6362, pp. 514–521. Springer, Heidelberg (2010)
3. Atasoy, S., Glocker, B., Giannarou, S., Mateus, D., Meining, A., Yang, G.Z., Navab, N.: Probabilistic Region Matching in Narrow-Band Endoscopy for Targeted Optical Biopsy. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009, Part I. LNCS, vol. 5761, pp. 499–506. Springer, Heidelberg (2009)
4. Atasoy, S., Mateus, D., Lallemand, J., Meining, A., Yang, G.Z., Navab, N.: Endoscopic Video Manifolds. In: Jiang, T., Navab, N., Pluim, J., Viergever, M.A. (eds.) MICCAI 2010, Part II. LNCS, vol. 6362, pp. 437–445. Springer, Heidelberg (2010)
5. Belkin, M., Niyogi, P.: Laplacian Eigenmaps for Dimensionality Reduction and Data Representation. *Neural Comput.* 15(6), 1373–1396 (2003)
6. Davies, D., Bouldin, D.: A Cluster Separation Measure. *IEEE Trans. on Pattern Anal.* (2), 224–227 (1979)
7. Figueiredo, M., Jain, A.: Unsupervised learning of finite mixture models. *IEEE Trans. on Pattern Anal.* 24(3), 381–396 (2002)
8. Hartigan, J., Wong, M.: A k-means clustering algorithm. *Journal of the Royal Statistical Society C* 28(1), 100–108 (1979)
9. Häussinger, K., Ballin, A., Becker, H., Bölskei, P., Dierkesmann, R., Dittrich, I., Frank, W., Freitag, L., Gottschall, R., Guschall, W., Hartmann, W., Hauck, R., Herth, F., Kirsten, D., Kohlhäuff, M., Kreuzer, A., Loddenkemper, R., Macha, N., Markus, A., Stanzel, F., Steffen, H., Wagner, M.: Recommendations for Quality Standards in Bronchoscopy. *Pneumologie* 58(5), 344 (2004)
10. He, X., Yan, S., Hu, Y., Niyogi, P., Zhang, H.: Face Recognition using Laplacian-faces. *IEEE Trans. on Pattern Anal.* 27(3), 328–340 (2005)
11. Mountney, P., Giannarou, S., Elson, D., Yang, G.-Z.: Optical Biopsy Mapping for Minimally Invasive Cancer Screening. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5761, pp. 483–490. Springer, Heidelberg (2009)

Focal Biologically Inspired Feature for Glaucoma Type Classification

Jun Cheng¹, Dacheng Tao², Jiang Liu¹, Damon Wing Kee Wong¹,
Beng Hai Lee¹, Mani Baskaran³, Tien Yin Wong^{3,4}, and Tin Aung^{3,4}

¹ Institute for Infocomm Research, Agency for Science,
Technology and Research, Singapore

{jcheng,jliu,wkwong,benghai}@i2r.a-star.edu.sg

² Centre for Quantum Computation and Intelligent Systems, FEIT,
University of Technology, Sydney, Australia

{dacheng.tao}@uts.edu.au

³ Singapore Eye Research Institute, Singapore

baskaran.mani@seri.com.sg, ophwty@nus.edu.sg, aung.tin@sneec.com.sg

⁴ Department of Ophthalmology, National University of Singapore, Singapore

Abstract. Glaucoma is an optic nerve disease resulting in loss of vision. There are two common types of glaucoma: open angle glaucoma and angle closure glaucoma. Glaucoma type classification is important in glaucoma diagnosis. Ophthalmologists examine the iridocorneal angle between iris and cornea to determine the glaucoma type. However, manual classification/grading of the iridocorneal angle images is subjective and time consuming. To save workload and facilitate large-scale clinical use, it is essential to determine glaucoma type automatically. In this paper, we propose to use focal biologically inspired feature for the classification. The iris surface is located to determine the focal region. The association between focal biologically inspired feature and angle grades is built. The experimental results show that the proposed method can correctly classify 85.2% images from open angle glaucoma and 84.3% images from angle closure glaucoma. The accuracy could be improved close to 90% with more images included in the training. The results show that the focal biologically inspired feature is effective for automatic glaucoma type classification. It can be used to reduce workload of ophthalmologists and diagnosis cost.

1 Introduction

Glaucoma is an optic nerve disease resulting in loss of vision. It is often associated with increased pressure of fluid inside the eye. Two common types of glaucoma are open angle glaucoma (OAG) and angle closure glaucoma (ACG). Ophthalmologists examine the iridocorneal angle between iris and cornea to determine OAG and ACG. When the angle is open, it is OAG. Otherwise, ACG. A detailed description of the angle structures can be found in [1]. Here we briefly explain why the iridocorneal angle is important. The iris, cornea, and lens are bathed in aqueous humor, which is continually produced by nearby tissues. It moves

out of the eye via the trabecular meshwork drainage. Blocking of the trabecular meshwork would lead to increased pressure in the eye. The trabecular meshwork is associated with the angle, thus, the iridocorneal angle is important. Because of different causes and specific treatments for different types of glaucoma as well as the necessity of urgent treatment of ACG, it is important to determine the glaucoma type early [2], which implies that it is essential to visualize the iridocorneal angle to make a correct diagnosis of the disease.

Gonioscopy is an eye examination looking at the front part of the eye between iris and cornea. It requires considerable clinical expertise and effort as well as a full knowledge of the angle structures [3]. Thus, it is not performed as often as it should be. A new option with much more convenience is the RetCam (Clarity Medical Systems, Inc., Pleasanton, CA) camera, which is explored to capture the image of iridocorneal angle [3] recently. Four typical iridocorneal angle images from inferior, superior, nasal and temporal quadrants of a left eye captured by RetCam are shown in Fig. 1. The angle which is of our interest is located at the boundary between the iris and the cornea. When other angle structures are visible on the cornea side of the iris surface, it is an open angle, otherwise, closed. Shaffer’s grading system [2] is widely used in gonioscopy to evaluate the angle status based on the visibility of the angle structures. In this paper, we focus on the clinically important grading: the classification between ACG and OAG. Manual grading of RetCam images is usually tedious, time consuming and expensive. Moreover, it is often subjective similar to many other medical applications [4] and thus reproducibility is a concern. To save workload and facilitate large-scale clinical use, it is essential to have a precise, efficient and cost effective system to determine glaucoma type automatically.

Automatic glaucoma type classification from iridocorneal images captured by RetCam is a new research topic and few work has been done for it. In [5], the

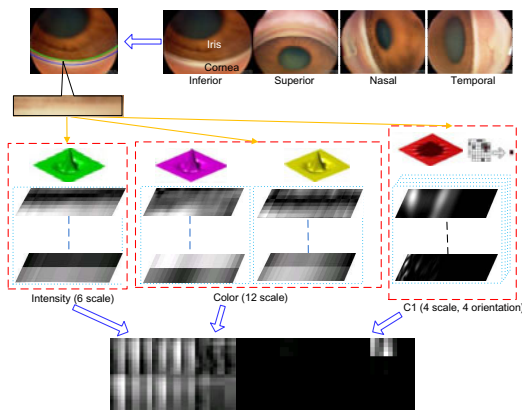


Fig. 1. Focal biologically inspired feature for glaucoma type classification: for each training image, we extract BIF from focal region. The feature consists of 6 intensity feature maps, 12 color feature maps, and 16 $C1$ units feature maps.

edges around strongest arc are used. The limitation is that some edges within iris are mistaken as edges from angle structures and reduce the classification accuracy. Moreover, it relies on the accuracy of edge detection as well as the sensitivities used to determine the edges. Based on the observation that human can capture the ‘gist’ of an image instantly, biologically inspired feature (BIF) [6][7][8] has been used in computer vision including scene classification, gait recognition and etc. BIF mimics the process of cortex for visual perception. Since the ophthalmologists classify the two types of glaucoma based on the visual information, we introduce BIF to aid them to reduce workload. Different from the application of scene classification where the scene varies largely from one to another, the ‘scene’ of an angle closure image differs slightly from that of an open one in a small region. When ophthalmologists examine the images, they focus on the region. Thus, it is essential to extract BIF from it, i.e., focal BIF. The proposed focal BIF as shown in Fig. 1 simulates the process.

The paper is organized as follows. In Section 1, we have given an introduction of the background and motivation for the system. Section 2 introduces the BIF. In Section 3, we introduce the methodology to determine the focal region and extract focal BIF for angle grading. Section 4 shows the experimental results, followed by the conclusions in the last section.

2 Biologically Inspired Feature

The BIF has proven to be effective in computer vision. The feature consists of 34 feature maps including 6 feature maps from intensity units, 12 feature maps from color units, and 16 feature maps from $C1$ units.

Among these features, the intensity units are obtained by convolving dyadic Gaussian pyramids with the intensity channel of a color image. The features correspond to the neurons of mammals which are sensitive to dark centers on bright surrounds or vice versa [9][10]. Nine spatial scales are generated with a ratio from 1:1 (level 0) to 1:256 (level 8). The intensity feature maps are obtained by the center-surround difference operation between center levels $c = 2, 3, 4$ and surround levels $s = c + d$, with $d = 3, 4$. Thus, six feature maps are computed at levels of 2-5, 2-6, 3-6, 3-7, 4-7, and 4-8. Because of the scale difference, maps of surround levels are interpolated to be the same size as the corresponding center levels, and then they are subtracted to generate the relevant feature maps, i.e., $I(c, s) = |I(c) - Interp_{s-c}I(s)|$.

The color units are inspired by the ‘color double-opponent’ system in cortex [9]. Neurons are excited by a color (e.g., blue) and inhibited by another color (e.g., yellow) in the center of receptive field, so are neurons in the surround. Herein, four color channels are used: $R = r - (g + b)/2$, $G = g - (r + b)/2$, $B = b - (r + g)/2$ and $Y = r + g - 2(|r - g| + b)$. For each color channel (R , G , B , and Y), dyadic Gaussian pyramids are used to generate nine spatial scales similar to intensity unit. Two color pairs $R - G$ and $B - Y$ are used. The feature maps are computed as the across scales center-surrounding differences. Similar to the computation of intensity units, surround maps are interpolated

to be the same size as the corresponding center maps and their difference is computed as: $RG(c, s) = |(R(c) - G(c)) - \text{Interp}_{s-c}(R(s) - G(s))|$ and $BY(c, s) = |(B(c) - Y(c)) - \text{Interp}_{s-c}(B(s) - Y(s))|$.

The $C1$ units are obtained by pooling over $S1$ units, which correspond to simple cells in $S1$ layer of the visual cortex. Gabor functions are used for feature extraction due to its similarity to the receptive field profiles in simple cells in $S1$ layer. The Gabor mother function is given by: $F(x, y) = \exp(-(x_0^2 + \gamma^2 y_0^2)/(2\delta^2)) \times \cos(2\pi x_0/\lambda)$, wherein $x_0 = x \cos \theta + y \sin \theta$, $y_0 = -x \sin \theta + y \cos \theta$, the range of x and y decides the scales of Gabor filters, and θ controls orientations. In this paper, eight scales with a range of sizes from 7×7 to 21×21 pixels with a step of two pixels are used. Four orientations are considered: 0° , 45° , 90° , and 135° . Thus, a total of 32 feature maps are obtained in $S1$ units. Pooling over two adjacent scales with an identical orientation, 16 feature maps are obtained from $C1$ units.

Arbitrarily extracting the BIF from the whole image does not work well for angle grading as the main difference between ACG and OAG lies in a small region, not to mention the difference due to various quadrants. In order to use BIF properly, we propose focal BIF. Focal BIF refers to biologically inspired feature from a focal region. In this application, it is the area between iris and cornea.

3 Methodology

A system for automatic glaucoma type classification is proposed with following steps: quadrant determination, focal region segmentation, and grading.

3.1 Quadrant Determination

As mentioned earlier, the images can be from inferior, superior, nasal and temporal quadrant of the eye. One important step for the automated diagnosis is to determine the quadrant. In this paper, quadrant is determined based on the location of the arc center and the location of edges. Canny edge [11] followed by circular Hough transform [12] as in [5] are used to obtain the strongest arc. Assuming $(x_i, y_i), i = 1, 2, \dots, N$ are the coordinates of all points from the arc inside the image, where top-left corner is defined as $(1, 1)$ and bottom-right corner as (m, n) , N is the number of points. The function to determine the quadrant Q is given as [1]:

$$Q = \begin{cases} \text{Superior, if } x_c - \bar{x}_i \geq |y_c - \bar{y}_i| \\ \text{Inferior, if } \bar{x}_i - x_c \geq |y_c - \bar{y}_i| \\ \text{Nasal, if } y_c - \bar{y}_i > |x_c - \bar{x}_i| \\ \text{Temporal, if } \bar{y}_i - y_c > |x_c - \bar{x}_i| \end{cases} \quad (1)$$

where $(\bar{x}_i, \bar{y}_i) = (\frac{1}{N} \sum_{i=1}^N x_i, \frac{1}{N} \sum_{i=1}^N y_i)$ is the mean of the edge coordinates, (x_c, y_c) is the center of the detected arc.

¹ A left eye is assumed here, swap nasal and temporal for a right eye.

3.2 Focal Region Segmentation

The focal region refers to the area human eye focused on when one examines the images. In this application, it is the area between iris and cornea. In order to extract focal region and align features from various regions, we propose to locate the iris surface as it is visible in both open angle and angle closure images. For angle closure, the iris surface is often the strongest ascending edge (from iris to cornea) in the focal region. However, for open angle, edges from other angle structures can be stronger. Thus, the output of the aforementioned circular Hough transform may find the iris surface inaccurately, as can be seen from the line in red in Fig. 2(a).

Without losing generality, assuming the image is from inferior side of an eye as in Fig. 2(a). Given $L_j(x) = I(x, j)$, $x = 1, 2, \dots, M$, from the j^{th} column of the image I . Assuming L_j crosses with the strongest arc at x_j . Inspired by the above observations on iris surface, we search for the point with strongest ascending edge (from iris to cornea) from pixels around x_j in L_i and get its coordinate x_k . Among all ascending Canny edge within $(x_k - w, x_k)$ as well as x_k itself, the point closest to pupil is used as candidate iris surface point in this column. Here, w is set to be estimated maximum angle width. Finally, curve fitting is applied based on all candidate points located in the last step. In this paper, the iris surface is modelled as part of circle and a circular Hough transform is applied again to find the fitted curve with circular center (x_c, y_c) and radius r . After obtaining the estimation of iris surface highlighted in green, another circular arc can be determined based same circular center (x_c, y_c) with a larger radius $r + \delta r$. The parameter δr is set to be slightly larger than w . The region in between is the focal region. As the side portions are often blurred, central portion would be used. Locating the iris surface is a critical step to find the focal region and then the focal BIF. Visually, the above algorithm finds the iris surface accurately in central portion for 393 of the 396 images. However, it is difficult to get a ground truth to compute a quantitative accuracy.

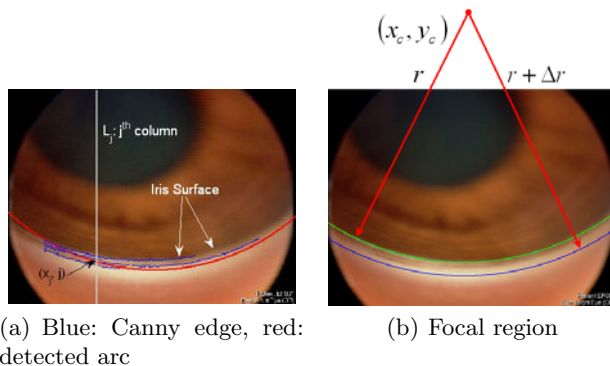


Fig. 2. Focal Region Segmentation

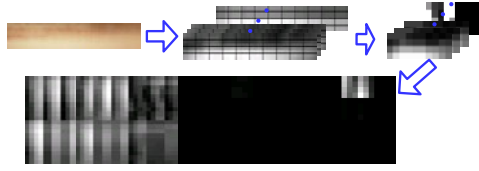


Fig. 3. Focal BIF from 4×10 sub-regions

A polar transform with respect to the circular center (x_c, y_c) as in Fig. 2(b) is then applied to turn the focal region into rectangular. The transform helps to align the focal region from images at different quadrants. As ophthalmologists examine only the central portion, only the feature maps from central portion of the focal region are used. Each feature map is divided into $m \times n$ sub-regions. The mean value of feature map in each sub-region is computed for final representation. Fig. 3 shows an example with feature maps divided into 4×10 sub-regions. Since each image is represented by 34 feature maps and each feature map is decomposed into $m \times n$ sub-regions, we have a total of $34mn$ mean values as the feature for each image.

3.3 Grading

In this paper, the grading problem is handled as a classification problem between ACG and OAG. Support vector machines (SVM) are used as the optimization tools for solving machine learning problems. The LIBSVM [13] is used in our experiments. RBF kernel is used with two-folder cross folder validation adopted to determine the parameters C and γ .

4 Experimental Results

A total of 99 different patient eyes as in [5] are used. For each patient eye, four quadrants are examined by an ophthalmologist, thus a total of $99 \times 4 = 396$ quadrants are evaluated and labelled as ACG or OAG quadrant. Among the gradings, 166 quadrants are with ACG and 230 quadrants are with OAG. These manual gradings are used as ground truth.

In the SVM training, 115 images from OAG quadrants together with same number of images from ACG quadrants are randomly selected for the training. The rest of images are used for testing. The SVM parameters C and γ are determined automatically through cross validation [13]. The training and testing are repeated five times to get an average result. Table 1 shows the percentage of corrected classified ACG and OAG by the proposed method for $m = 4, n = 10$ together the results by prior work [5] as well as the results when other classifiers are used. The results show conventional neural networks such as multi-layer perception (MLP), and k-nearest-neighbours (kNN) cannot perform as well as SVM on this task. To show the effectiveness of BIF, we also conduct tests by replacing the BIF with simplified features such as color histogram (CH), color

Table 1. Performance of the proposed method and other methods

Method	Prior [5]	MLP	kNN	Proposed
OAG	80.3%	77.4%	51.8%	85.2%
ACG	80.3%	74.3%	65.0%	84.3%

Table 2. Performance using other features

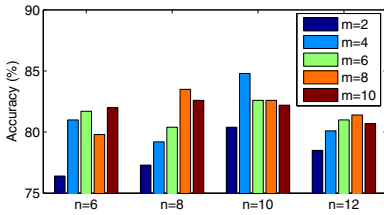
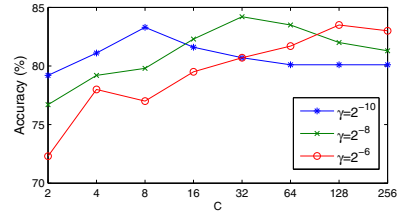
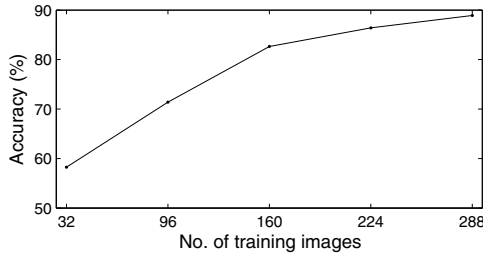
Method	CH	CM	IH
OAG	59.0%	65.1%	54.2%
ACG	67.5%	57.8%	56.6%

moment (CM) and intensity histogram (IH). The results summarized in Table 2 show that these features are not suitable for this task.

We also conduct tests with other m and n combinations. The classification accuracy is computed as the average accuracy of ACG and OAG and summarized in Fig. 4. The results show a slight performance drop with other settings. The average accuracy increases as n increases to 10. For larger n , it drops as too much redundant information leads to a biased classifier. A similar phenomena is observed as m changes.

The two SVM parameters C and γ are important as well. Although C and γ are selected automatically, it is still necessary to look at the results with other parameters. Fig. 5 shows the performance with different C and γ combinations from exponentially growing C and γ for the case $m = 4$ and $n = 10$.

The number of training samples is another critical factor. Fig. 6 shows the results when different number of training images are used in the training. The results show that the accuracy can be improved close to 90% with more images included in the training.

**Fig. 4.** Performance when each feature map is divided into $m \times n$ sub-regions**Fig. 5.** Performance with various C and γ **Fig. 6.** Performance with different number of training images

5 Conclusion

Glaucoma type classification is important in glaucoma diagnosis. Clinically, automatic glaucoma type classification helps reduce workload of ophthalmologists and diagnosis cost. It make it possible for large-scale clinical use and benefit patients. However, it is a challenging work due to ambiguous angle structures in some images. In this paper, we explore focal BIF for the classification. The association between focal BIF and glaucoma type is built through SVM learning. The results show that focal BIF is effective for the classification with 85.2% OAG and 84.3% ACG correctly detected based on 4×10 sub-regions. With more images included in the training, the accuracy can be improved close to 90%. A limitation is that the ground truth used in this paper is from one ophthalmologist only and thus can be biased. In the future, gradings from more ophthalmologists are to be used. Future work would focus on feature representation to further improve the accuracy, with more images from different patients.

References

1. Minckler, D., Foster, P., Hung, P.T.: Angle Closure Glaucoma Classification and Racial Variation. *Asia Journal of Ophthalmology* 3(3,4), 3–9 (2001)
2. Salmon, J.F.: The Diagnostic Importance of Gonioscopy, <http://www.glaucomaworld.net/english/019/e019a01t.html/>
3. Ahmed, I.I.K., Mackeen, L.D.: A New Approach for Imaging the Angle. *Glaucoma Today*, 27–30 (2007)
4. Huang, W., Li, H., Chan, K.L., Lim, J.H., Liu, J., Wong, T.Y.: A Computer Aided Diagnosis System of Nuclear Cataract via Ranking. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) *MICCAI 2009*. LNCS, vol. 5762, pp. 803–810. Springer, Heidelberg (2009)
5. Cheng, J., Liu, J., Lee, B. H., Wong, D., Yin, F., Aung, T., Baskaran, M., Perera, S., Wong, T.Y.: Closed Angle Glaucoma Detection in Retcam Images. In: *Proc. of Int. Conf. of the IEEE Engineering in Med. and Bio. Society*, pp. 4096–4099 (2010)
6. Siagian, C., Itti, L.: Rapid Biologically-inspired Scene Classification using Features Shared with Visual Attention. *IEEE Trans. Pattern Anal. Mach. Intell.* 31(2), 260–274 (2008)
7. Song, D., Tao, D.: Biologically Inspired Feature Manifold for Scene Classification. *IEEE Trans. Image Processing* 19(1), 174–184 (2010)
8. Mu, Y., Tao, D.: Biologically Inspired Feature Manifold for Gait Recongition. *Neurocomputing* 73, 895–902 (2010)
9. Itti, L., Koch, C., Niebur, E.: A Model of Saliency-based Visual Attention for Rapid Scene Analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* 20(11), 1254–1259 (1998)
10. Leventhal, A.G.: *The Neural Nasis of Visual Function*. Vision and Visual Dysfunction (1991)
11. Canny, J.: A Computational Approach to Edge Detection. *IEEE Trans. Pattern Anal. and Mach. Intell.* 8, 679–698 (1986)
12. Morse, B.S.: *Segmentation (Edge based, Hough Transform)*. Brigham Young University, Lecture Notes (2000)
13. Chang, C.C., Lin, C.J.: *LIBSVM: A Library for Support Vector Machines*, <http://www.csie.ntu.edu.tw/~cjlin/libsvm>

The Relevance Voxel Machine (RVoxM): A Bayesian Method for Image-Based Prediction

Mert R. Sabuncu^{1,2} and Koen Van Leemput^{1,2,3,4}

¹ Athinoula A. Martinos Center for Biomedical Imaging, Department of Radiology, MGH, Harvard Medical School, USA

² Computer Science and Artificial Intelligence Laboratory, MIT, USA

³ Departments of Information and Computer Science and of Biomedical Engineering and Computational Science, Aalto University, Finland

⁴ Department of Informatics and Mathematical Modeling, Technical University of Denmark, Denmark

Abstract. This paper presents the Relevance Voxel Machine (RVoxM), a Bayesian multivariate pattern analysis (MVPA) algorithm that is specifically designed for making predictions based on image data. In contrast to generic MVPA algorithms that have often been used for this purpose, the method is designed to utilize a small number of spatially clustered sets of voxels that are particularly suited for clinical interpretation. RVoxM automatically tunes all its free parameters during the training phase, and offers the additional advantage of producing probabilistic prediction outcomes. Experiments on age prediction from structural brain MRI indicate that RVoxM yields biologically meaningful models that provide excellent predictive accuracy.

Keywords: Multivariate Pattern Analysis, MRI.

1 Introduction

Medical imaging commonly entails relating image content to a clinical or experimental condition. Traditional univariate approaches, such as voxel-based morphometry [2], generate anatomical maps of the effects by analyzing each location individually. MVPA methods, in contrast, offer increased specificity and sensitivity for predicting the outcome by considering all voxels *simultaneously* [6,8,10,12,14,16,17,15]. Yet studies on image-based prediction have typically employed generic MVPA methods, such as Support or Relevance Vector Machines (SVMs/RVMs) [5,18], which do not account for the spatial organization of *imaging* data.

As demonstrated in semi-supervised learning, significant performance gains can be obtained by explicitly utilizing the underlying structure of the data [3,4]. One approach to achieve this with images is to impose an a priori model on the covariation of voxel measurements – a strategy that has proven powerful in computer vision [13]. Further motivation for such image-based prediction models is *interpretability*: rather than a “black box” tool, we are also interested in understanding and visualizing the key areas that drive predictions. Although it is

possible to display the workings of generic linear MVPA methods as images [12], the results are often scattered and hard to interpret biologically [7].

In this paper, we present the Relevance Voxel Machine (RVoxM), a novel MVPA algorithm that is specifically designed for image-based prediction. It uses a Bayesian approach and builds largely on existing RVM machinery to obtain not only good prediction performance, but also sparse solutions. Unlike RVMs, however, where sparseness is realized by discarding many of the samples, i.e., *training subjects*, our approach removes most *voxels*, retaining only those voxels that are relevant for prediction. Furthermore, our model encourages spatial clustering of these “relevance voxels” and computes predictions as linear combinations of their content, yielding results that are both biologically plausible and intuitive to interpret. Compared to related efforts that incorporate spatial context within the SVM or penalized regression frameworks [15,7], our method inherits all the usual advantages of RVMs over non-Bayesian methods, including providing probabilistic outcomes and the automatic tuning of all free parameters.

We test RVoxM on the problem of estimating the age of healthy subjects from structural brain MRI scans, and show that it achieves high accuracy using a pattern of “relevance voxels” that easily lends itself to biological interpretation.

2 Model

We use a generative model similar to the one of RVM [18]. Let t denote a real-valued *target* variable (e.g., age) that we aim to predict from image data, and x_i a voxel-level measurement (e.g., gray matter density) at the voxel indexed by i . We define a Gaussian conditional distribution for t : $p(t|\mathbf{x}, \mathbf{w}, \beta) = \mathcal{N}(t|y(\mathbf{x}), \beta^{-1})$, with variance β^{-1} and a mean that is given by the linear model

$$y(\mathbf{x}) = \sum_{i=1}^M x_i w_i + w_0 = \mathbf{w}^T \mathbf{x}, \quad (1)$$

where $\mathbf{w} = (w_0 \cdots w_M)^T$ are adjustable “weights” encoding the strength of each voxel’s contribution to the prediction, $\mathbf{x} = (1, x_1, \cdots, x_M)^T$ denotes the vectorized image the prediction is based on, and M is the number of voxels. For notational convenience, we include an extra “voxel” to account for the bias, w_0 .

We assume a zero-mean Gaussian prior distribution over \mathbf{w} :

$$p(\mathbf{w}|\boldsymbol{\alpha}, \lambda) = \mathcal{N}(\mathbf{w}|\mathbf{0}, \mathbf{P}^{-1}),$$

where \mathbf{P} is a $(M+1) \times (M+1)$ precision (inverse covariance) matrix defined as

$$\mathbf{P} = \text{diag}(\alpha_0, \cdots, \alpha_M) + \lambda \mathbf{K}.$$

Here, $\boldsymbol{\alpha} = (\alpha_0, \cdots, \alpha_M)^T$ and λ are *hyperparameters*, and \mathbf{K} is a fixed, positive-semidefinite matrix that encourages local spatial smoothness of \mathbf{w} . In particular, we use $\mathbf{K} = \boldsymbol{\Upsilon}^T \boldsymbol{\Upsilon}$, where $\boldsymbol{\Upsilon}$ is a sparse matrix in which each row corresponds to a pair of neighboring voxels in the image. For neighboring voxels $\{i, j\}$, the corresponding row has zero entries everywhere except for the i^{th} and j^{th} column, which have entries -1 and 1 , respectively. Re-writing the prior as

$$p(\mathbf{w}|\boldsymbol{\alpha}, \lambda) \propto \exp\left(-\frac{1}{2}\sum_{i=0}^M \alpha_i w_i^2 - \frac{\lambda}{2}\|\boldsymbol{\Upsilon}\mathbf{w}\|^2\right)$$

shows that it encodes a preference for models that are both *sparse and spatially clustered*: we explicitly seek models that explain t through a small collection of image patches that easily lend themselves to neuroscientific interpretation. The fact that there is a hyperparameter α_i associated with each voxel’s weight w_i is responsible for achieving sparsity in those weights – in practice many of the α_i ’s tend to very large values, forcing the corresponding weights to zero and “switching off” the contribution of many voxels. Importantly, we also explicitly take the spatial structure of image data into account by penalizing large entries in the vector $\boldsymbol{\Upsilon}\mathbf{w}$, which represent large differences between the weights of neighboring voxels. Thus, we encode a preference for spatial clusters of “switched-on” voxels, as these are both biologically more plausible and easier to interpret than speckles of isolated voxels scattered throughout the image area.

3 Hyperparameter Estimation

Given training data, $\{\mathbf{x}_n, t_n\}_{n=1}^N$, where \mathbf{x}_n is the n^{th} training image and t_n its target variable, our first goal is to determine the values of the hyperparameters $\boldsymbol{\alpha}$, λ , and β . Using type-II maximum likelihood, we estimate the hyperparameters by maximizing the marginal likelihood function obtained by integrating out \mathbf{w} :

$$\begin{aligned} p(\mathbf{t}|\mathbf{X}, \boldsymbol{\alpha}, \lambda, \beta) &= \int_{\mathbf{w}} \left(\prod_{n=1}^N p(t_n|\mathbf{x}_n, \mathbf{w}, \beta) \right) p(\mathbf{w}|\boldsymbol{\alpha}, \lambda) d\mathbf{w} \\ &= \int_{\mathbf{w}} \left(\frac{\beta}{2\pi} \right)^{N/2} \exp\left(-\frac{\beta}{2}\|\mathbf{t} - \mathbf{X}\mathbf{w}\|^2\right) \frac{|\mathbf{P}|^{1/2}}{(2\pi)^{M/2}} \exp\left(-\frac{1}{2}\mathbf{w}^T\mathbf{P}\mathbf{w}\right) d\mathbf{w} \\ &= \frac{|\boldsymbol{\Gamma}|^{-1/2}}{(2\pi)^{N/2}} \exp\left(-\frac{1}{2}\mathbf{t}^T\boldsymbol{\Gamma}^{-1}\mathbf{t}\right), \end{aligned} \quad (2)$$

where $\mathbf{t} = (t_1, \dots, t_N)^T$, $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N]^T$ is the $N \times (M+1)$ “design” matrix, and we have defined the $N \times N$ matrix $\boldsymbol{\Gamma}$ given by

$$\boldsymbol{\Gamma} = \beta^{-1}\mathbf{I} + \mathbf{X}\mathbf{P}^{-1}\mathbf{X}^T.$$

We take a “coordinate-ascent” approach to maximize Eq. (2). We first define:

$$\boldsymbol{\mu} = \beta\boldsymbol{\Sigma}\mathbf{X}^T\mathbf{y}, \quad \boldsymbol{\Sigma} = (\beta\mathbf{X}^T\mathbf{X} + \mathbf{P})^{-1}. \quad (3)$$

Fixing λ , β , and $\{\alpha_j\}$ for all $j \neq i$, differentiating the log of Eq. (2) w.r.t α_i , equating to zero and rearranging yields the following update:

$$\alpha_i^{\text{new}} = \frac{\gamma_i}{\mu_i^2} \alpha_i, \quad (4)$$

where $\gamma_i = 1 - \alpha_i\boldsymbol{\Sigma}_{ii} - \lambda(\mathbf{P}^{-1}\mathbf{K})_{ii}$. Similarly, fixing $\boldsymbol{\alpha}$ and β , differentiating w.r.t λ , and rearranging yields the following update equation for λ :

$$\lambda^{new} = \frac{\text{trace}(\Delta \mathbf{K})}{\boldsymbol{\mu}' \mathbf{K} \boldsymbol{\mu}} \lambda, \quad (5)$$

where $\Delta = \mathbf{P}^{-1} - \Sigma$. Similarly, an update for β can be derived:

$$\beta^{new} = \frac{N}{\|\mathbf{y} - \mathbf{X}\boldsymbol{\mu}\|^2 + \text{trace}(\mathbf{X}\Sigma\mathbf{X}^T)}. \quad (6)$$

Optimization now proceeds by cycling through these equations. We initialize with $\alpha_i = 1, \forall i$, $\lambda = 1$ and $\beta = 0$; monitor the objective function at each iteration and terminate when the increase over the previous iteration is below a tolerance. Although currently we have no theoretical guarantees that the presented equations indeed increase the objective function, we have not encountered any situation where this was not the case in our experiments.

4 The RVoxM Learning Algorithm

In practice, most ($> 90\%$) α_i 's tend to grow to infinity, effectively clamping the corresponding weight w_i 's to zero and removing those voxels from the model. We exploit this to obtain a greedy learning algorithm for large 3-D image volumes, using two computational tricks. First, each time one of the α_i 's exceeds a certain (very large) value, the corresponding voxel is pruned from the model and computations continue based on remaining voxels only, in a manner similar to the RVM algorithm [18]. Second, we use a multi-resolution approach commonly employed in image processing. We construct a pyramid representation of the training images, where each level consists of lower-resolution images computed by subsampling the images from the previous resolution. The algorithm then starts by learning the hyperparameters for the lowest resolution images, propagates them down for the initialization of the next level, and so forth until the final resolution level is reached; voxels that were pruned at the previous level remain so henceforth. Although this greedy algorithm prevents voxels from re-entering once they have been removed, our experiments suggest that it works quite well in practice.

5 Using RVoxM to Make Predictions

Having learned the hyperparameters $\boldsymbol{\alpha}^*$, λ^* , and β^* from the training data, we can make predictions about t for a new input image \mathbf{x} by evaluating the posterior

$$p(t|\mathbf{x}, \mathbf{X}, \mathbf{t}, \boldsymbol{\alpha}^*, \lambda^*, \beta^*) = \int_{\mathbf{w}} p(t|\mathbf{x}, \mathbf{w}, \beta^*) p(\mathbf{w}|\mathbf{X}, \mathbf{t}, \boldsymbol{\alpha}^*, \lambda^*) d\mathbf{w}.$$

It can be shown that this distribution is a Gaussian with mean

$$\boldsymbol{\mu}^T \mathbf{x} \quad (7)$$

and variance $\frac{1}{\beta^*} + \mathbf{x}^T \Sigma \mathbf{x}$, where $\boldsymbol{\mu}$ and Σ are given by Eq. (3) in which $\boldsymbol{\alpha}$, λ , and β have been set to their optimized values $\boldsymbol{\alpha}^*$, λ^* , and β^* .

In the remainder, we will use the maximum a posteriori value given by Eq. (7) to predict t , which corresponds to the linear model of Eq. (11) in which the voxels' weights \mathbf{w} are set to $\boldsymbol{\mu}$. In many voxels $\mu_i = 0$ (because their α_i was set to infinity) – we call the remaining voxels the “relevance voxels” as these are the only ones effectively used to predict the target variable t .

6 Experimental Results

We applied RVoxM to the problem of estimating a person's age from a brain MRI scan. This problem has attracted recent attention [19, 11] since it provides a novel perspective for studying healthy development and aging patterns, while characterizing pathologic deviations in disease.

We used T1-weighted scans from 336 cognitively normal subjects (age range 18-93 years), available through the OASIS dataset¹. We processed all scans with SPM8², using default settings, to obtain spatially aligned gray matter maps. The *gray matter density* values (tissue probabilities modulated by the Jacobian of the non-linear warp) were used as voxel-level measurements x_i in the experiment. To assess generalization accuracy, we split the data into two arbitrary³ halves

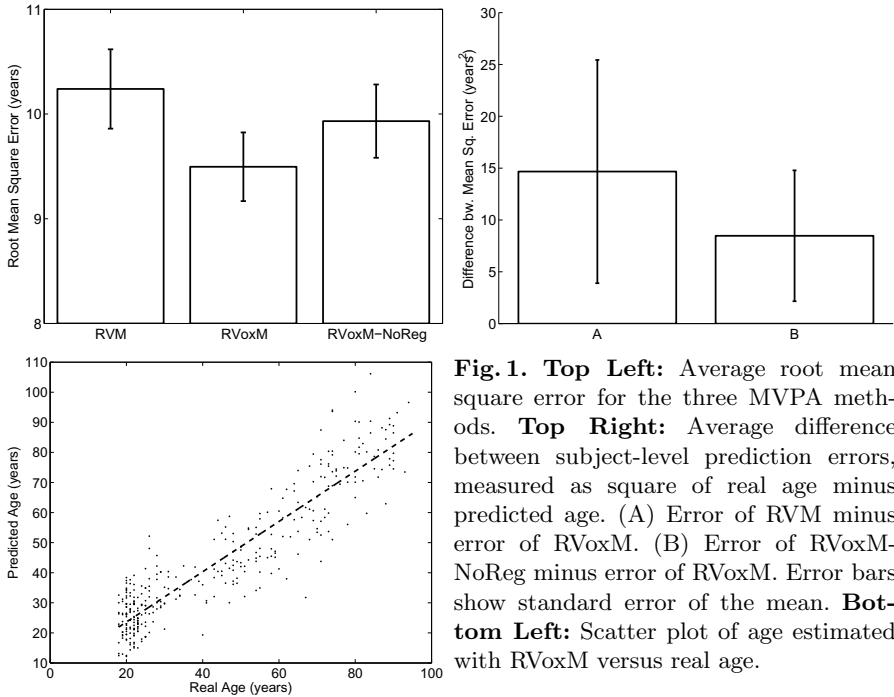


Fig. 1. Top Left: Average root mean square error for the three MVPA methods. **Top Right:** Average MVPA difference between subject-level prediction errors, measured as square of real age minus predicted age. (A) Error of RVM minus error of RVoxM. (B) Error of RVoxM-NoReg minus error of RVoxM. Error bars show standard error of the mean. **Bottom Left:** Scatter plot of age estimated with RVoxM versus real age.

¹ <http://www.oasis-brains.org>. 1.5T Siemens Vision scanner, $1 \times 1.25 \text{mm}^3$, MPRAGE.

² <http://www.fil.ion.ucl.ac.uk/spm/software/spm8/>

³ Simply based on the alphabetical ordering of the anonymized filenames.

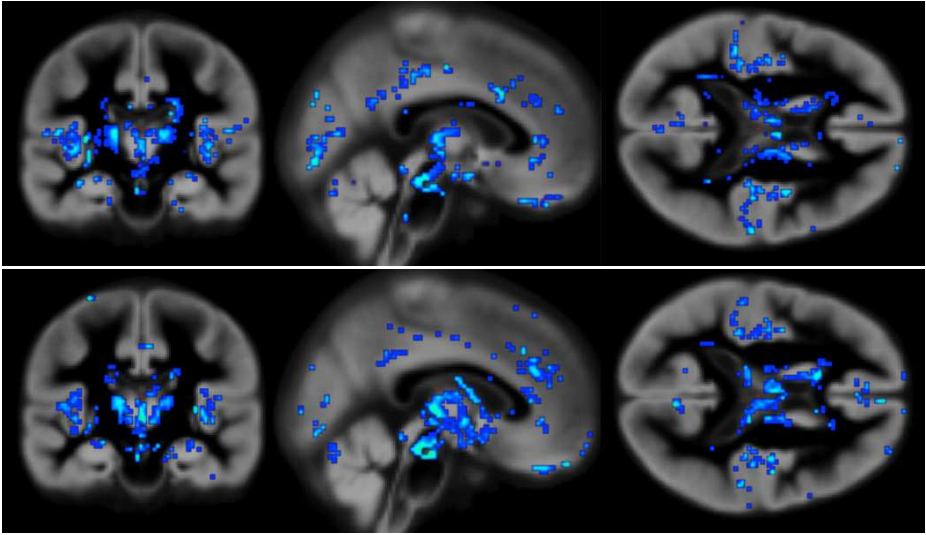


Fig. 2. Relevance voxels (in blue) for predicting age, overlaid on the average gray matter density image across all subjects. Brighter blue indicates a higher absolute value, and thus a higher relevance for prediction. Top row: Model from training on the first half of the data. Bottom row: Model from training on the second half of the data.

(age and sex matched, 43.7 ± 23.8 years, 62.5% female). We employed each group to train the RVoxM, which was then applied to the complementary group for testing. All reported results are averages across the two training/testing sessions.

In addition to RVoxM, we used two other methods as benchmarks. The first method (“RVM”) is another approach for estimating age from structural MRI [11]. It uses a principal component analysis to achieve a dimensionality-reduced representation of the images, and subsequently applies a linear RVM algorithm. We used the optimal implementation settings described in [11] and a public implementation of RVM⁴. The second benchmark (“RVoxM-NoReg”) was an implementation of RVoxM with no spatial regularization, i.e., with the hyperparameter λ clamped to zero. A comparison with the latter benchmark gives us an insight into the effect of spatial regularization on the results.

Fig. 1 (top left) illustrates the root mean square error (RMSE) for the three algorithms. On average, RVoxM yields the best accuracy with a RMSE less than 9.5 years (paired t-test, $P < 0.05$); Fig. 1 (bottom left) plots the age predicted by RVoxM for each subject versus the subject’s real age. Fig. 1 (top right) plots the average difference between the individual-level prediction errors (square of predicted age minus true age) obtained by RVoxM and the other two methods. On average, RVoxM achieves a statistically significantly smaller prediction error at the individual-level. RVoxM also attains the highest correlation (r-value)

⁴ <http://www.vectoranomaly.com/downloads/downloads.htm>

between the subjects’ real age and predicted age among all three methods: 0.92 for RVoxM vs. 0.90, and 0.91 for RVM and RVoxM-NoReg, respectively⁵.

Fig. 2 shows μ , RVoxM’s estimated voxel weights, for each of the two training sessions. Recalling that the prediction on new data is simply the linear product between μ and the test image (Eq. (7)), the value of μ at a specific voxel reflects the contribution of that voxel to the prediction. It can be appreciated that most voxels have a zero contribution (i.e., the model is sparse), and that the “relevance voxels” (with a non-zero contribution) occur in clusters, providing clear clues as to what parts of the gray matter are driving the age prediction process. Furthermore, the relevance voxels exhibit an overall very similar pattern across the two training sessions, providing evidence that these patterns are likely to be associated with the underlying biology and can be interpreted. We leave the interpretation of these relevance voxel patterns to future work.

7 Conclusion

In this paper, we proposed a novel Bayesian framework for image-based prediction. The proposed method yields a model where the predicted outcome is a linear combination of a small number of spatially clustered sets of voxels. We developed a computationally efficient optimization algorithm, RVoxM, to learn the properties of this model from a training data set. While RVoxM is not guaranteed to find the global optimum, our empirical results suggest that it finds a good solution in practice. Experiments on age prediction from structural brain MRI indicate that RVoxM derives excellent predictive accuracy from a small pattern of voxels that easily lends itself to neuroscientific interpretation.

Although we have used a regression model in this paper, it is straightforward to extend the technique to probabilistic classification by introducing a logistic sigmoid function [18]. In future work, we thus intend to apply RVoxM to also predict dichotomous outcomes (e.g., diagnosis), in addition to continuous ones.

Acknowledgments. Support for this research was provided by the NIH NCRR (P41-RR14075), the NIBIB (R01EB006758, R01EB013565), the NINDS (R01NS052585), the Academy of Finland (133611), the Finnish Funding Agency for Technology and Innovation (ComBrain), and MGH ECOR (2010A055663). MS received support from a KL2 Medical Research Investigator Training grant awarded via Harvard Catalyst (NIH grant 1KL2RR025757-01 and financial contributions from Harvard and its affiliations).

References

1. Ashburner, J.: A fast diffeomorphic image registration algorithm. *NeuroImage* 38(1), 95–113 (2007)
2. Ashburner, J., Friston, K.: Voxel-based morphometry—the methods. *NeuroImage* 11(6), 805–821 (2000)

⁵ We note that [11] reported slightly better correlation values for RVM ($r = 0.92$), which is probably due to the increased sample size ($N = 550$) and/or different data.

3. Batmanghelich, N., Taskar, B., Davatzikos, C.: A general and unifying framework for feature construction, in image-based pattern classification. In: Prince, J., Pham, D., Myers, K. (eds.) IPMI 2009. LNCS, vol. 5636, pp. 423–434. Springer, Heidelberg (2009)
4. Belkin, M., Niyogi, P., Sindhvani, V.: Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. *The Journal of Machine Learning Research* 7, 2399–2434 (2006)
5. Cortes, C., Vapnik, V.: Support-vector networks. *Machine Learning* 20(3), 273–297 (1995)
6. Cox, D., Savoy, R.: Functional magnetic resonance imaging (fMRI) “brain reading”: detecting and classifying distributed patterns of fMRI activity in human visual cortex. *NeuroImage* 19(2), 261–270 (2003)
7. Cuingnet, R., Chupin, M., Benali, H., Colliot, O.: Spatial prior in SVM-based classification of brain images. In: *Proceedings of SPIE*, vol. 7624, p. 76241L (2010)
8. Davatzikos, C., Fan, Y., Wu, X., Shen, D., Resnick, S.: Detection of prodromal Alzheimer’s disease via pattern classification of magnetic resonance imaging. *Neurobiology of Aging* 29(4), 514–523 (2008)
9. Dosenbach, N., Nardos, B., Cohen, A., Fair, D., Power, J., et al.: Prediction of Individual Brain Maturity Using fMRI. *Science* 329(5997), 1358 (2010)
10. Fan, Y., Batmanghelich, N., Clark, C., Davatzikos, C.: Spatial patterns of brain atrophy in MCI patients, identified via high-dimensional pattern classification, predict subsequent cognitive decline. *NeuroImage* 39(4), 1731–1743 (2008)
11. Franke, K., Ziegler, G., Kloppel, S., Gaser, C.: Estimating the age of healthy subjects from T1-weighted MRI scans using kernel methods: Exploring the influence of various parameters. *NeuroImage* 50(3), 883–892 (2010)
12. Kloppel, S., Stonnington, C., Chu, C., Draganski, B., Scahill, R., et al.: Automatic classification of MR scans in Alzheimer’s disease. *Brain* 131(3), 681 (2008)
13. Li, S.: *Markov random field modeling in image analysis*. Springer-Verlag New York Inc., New York (2009)
14. Magnin, B., Mesrob, L., Kinkingnéhun, S., Péligrini-Issac, M., Colliot, O., et al.: Support vector machine-based classification of Alzheimer’s disease from whole-brain anatomical MRI. *Neuroradiology* 51(2), 73–83 (2009)
15. Michel, V., Gramfort, A., Varoquaux, G., Eger, E., Thirion, B.: Total variation regularization for fMRI-based prediction of behaviour. *Arxiv preprint arXiv:1102.1101* (2011)
16. Pereira, F., Mitchell, T., Botvinick, M.: Machine learning classifiers and fMRI: a tutorial overview. *NeuroImage* 45(1), S199–S209 (2009)
17. Pohl, K., Sabuncu, M.: A unified framework for MR based disease classification. In: Prince, J., Pham, D., Myers, K. (eds.) IPMI 2009. LNCS, vol. 5636, pp. 300–313. Springer, Heidelberg (2009)
18. Tipping, M.: Sparse Bayesian learning and the relevance vector machine. *Journal of Machine Learning Research* 1, 211–244 (2001)

Assessment of Regional Myocardial Function via Statistical Features in MR Images

Mariam Afshin^{1,3}, Ismail Ben Ayed², Kumaradevan Punithakumar²,
Max W.K. Law^{1,2}, Ali Islam⁴, Aashish Goela⁵, Ian Ross⁵,
Terry Peters^{1,3}, and Shuo Li^{1,2}

¹ University of Western Ontario, London, ON, Canada

² GE Healthcare, London, ON, Canada

³ Robarts Research Institute, London, ON, Canada

⁴ St. Joseph's Health Care, London, ON, Canada

⁵ London Health Science Center, London, Ontario, Canada

Abstract. Early and accurate detection of Left Ventricle (LV) regional wall motion abnormalities significantly helps in the diagnosis and follow-up of cardiovascular diseases. We present a regional myocardial abnormality detection framework based on image statistics. The proposed framework requires a minimal user interaction, only to specify initial delineation and anatomical landmarks on the first frame. Then, approximations of regional myocardial segments in subsequent frames were systematically obtained by superimposing the initial delineation on the rest of the frames. The proposed method exploits the Bhattacharyya coefficient to measure the similarity between the image distribution within each segment approximation and the distribution of the corresponding user-provided segment. Linear Discriminate Analysis (LDA) is applied to find the optimal direction along which the projected features are the most descriptive. Then a Linear Support Vector Machine (SVM) classifier is employed for each of the regional myocardial segments to automatically detect abnormally contracting regions of the myocardium. Based on a clinical dataset of 30 subjects, the evaluation demonstrates that the proposed method can be used as a promising diagnostic support tool to assist clinicians.

1 Introduction

Heart failure is a prevalent disease that can be caused by various heart conditions, in particular, ischemic heart disease (IHD) [1]. The decrease of blood supply produced by coronary artery stenosis impairs the contractile properties of specific myocardial areas. This deviates the normal regional wall motion and contractility patterns of the myocardium, especially the left ventricle (LV). Early and accurate detection of LV regional wall motion abnormalities significantly helps in the diagnosis and follow-up of IHD [2]. In routine clinical use, cardiac function is estimated by visual assessment and interpretation of LV and, therefore, it is highly subject-dependent. Alternatively, computer-aided detection

systems have been attempted in recent years in order to automatically analyze LV myocardial function quantitatively [3], and to classify hearts into normal or abnormal groups [4]. In clinical practice, the regional myocardial function is commonly scored by following American Heart Association (AHA) standards [5], where the LV is divided into 17 segments. Existing regional heart function analysis methods are based on information theoretic measures and unscented Kalman filter approaches [6], differentiable manifolds [7], independent component analysis classifier [8], pattern recognition technique based on intra-segment correlation [9], and tensor-based classification [10]. Most of the existing methods require extensive user interaction or computationally expensive segmentation algorithms. This study investigates assessment of regional myocardial function using MR statistics and starting from a minimal user input. Typically cardiac MR data consist of 10 sequences, each comprising 20 or 25 temporal image frames. From a simple user input, we computed image statistics that are related to myocardium function. Given a user-provided delineation of the first frame, approximations of regional myocardial segments in subsequent frames were systematically obtained by superimposing the initial delineation on the rest of the frames. The proposed method exploits the Bhattacharyya coefficient [11] to measure the similarity between the image distribution within each segment approximation and the distribution of the corresponding user-provided segment. Linear Discriminate Analysis (LDA) is applied to find the optimal direction along which the projected features are the most descriptive. Linear Support Vector Machine (SVM) classifier is then employed for each of the regional myocardial segments to automatically detect abnormal functional regions of the myocardium. The proposed method performs significantly better than other recent methods and requires fewer computational resources. The evaluations performed on a clinical dataset of 30 subjects show that the proposed method is a promising diagnostic support tool.

2 Constructing Image Statistics

We consider image statistics as representative features to classify regional myocardium into normal or abnormal classes. Let \mathcal{I} be a set of cardiac MR images of a single slice containing N frames [1]. Let \mathbf{I} be a reference image which is an end-diastolic frame corresponding to the largest volume during cardiac cycle, whose delineation is given as depicted in Fig. 1(a). Let $\Gamma_{in}, \Gamma_{out} : [0, 1] \rightarrow \Omega$ denote respectively the corresponding manual endo and epi-cardial boundaries of \mathbf{I} . We divide \mathbf{I} into M regional segments [2] following the AHA standard [5], and using *anatomical landmarks* [3]. Fig. 1(b) shows the regional segments for \mathbf{I} . For $(n, m) \in [1 \dots N] \times [1 \dots M]$, let I_{nm} denotes the regional cardiac segment corresponding to segment m in frame n , and Γ_{out}^{nm} the boundary of I_{nm} (refer

¹ The number of frames N is typically equal to 20 or 25.

² M would be 4, 6, and 6 for apical, mid-cavity and basal slices, respectively.

³ As suggested by [5] the attachment of the right ventricular wall to the LV is used to identify and separate the septum from the LV anterior and inferior free walls.

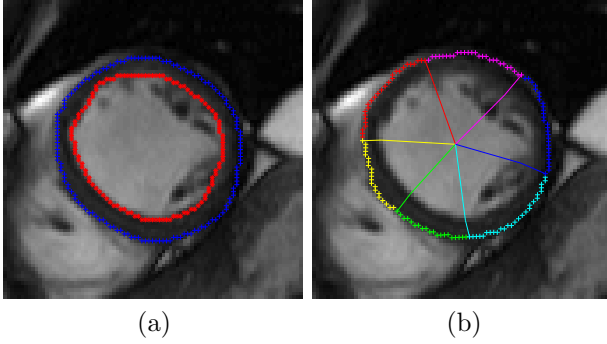


Fig. 1. (a) Manual delineation of the reference image \mathbf{I} for a mid-cavity slice. (b) Regional segments of the reference image \mathbf{I} for mid-cavity slice.

to Fig. 2 (a)). The classification procedures are identical for apical, mid-cavity and basal slices. Let us now superimpose the region defined by the epi-cardial boundary Γ_{out}^{nm} to the other frames in the sequence as shown in Fig. 2(a-c), and compute the corresponding image statistics (Fig. 2(d-f)). We define $\mathbf{R}_\Gamma \subset \Omega$ to be the region enclosed within $\Gamma \in \{\Gamma_{in}, \Gamma_{out}, \Gamma_{out}^{nm}\}$, and $P_{\mathbf{R}_\Gamma, \mathbf{I}}$ the intensity distribution of \mathbf{I} within region \mathbf{R}_Γ :

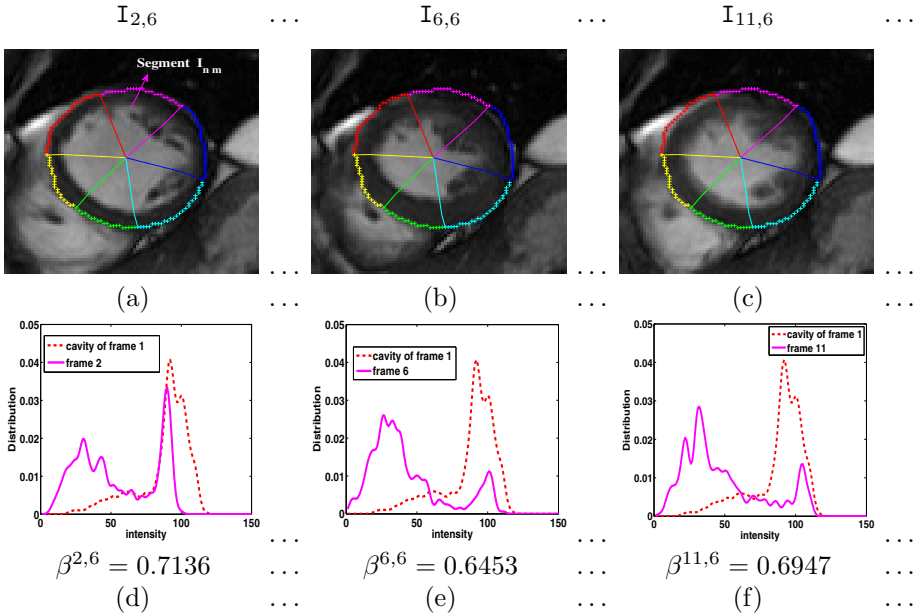


Fig. 2. (a-c): Regional myocardial segments superimposed on subsequent frames. (d-f): the corresponding image statistics.

$$P_{\mathbf{R}_r, \mathbf{I}}(z) = \frac{\int_{\mathbf{R}_r} K(z - \mathbf{I}) dx}{a_{\mathbf{R}_r}}, \quad K(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp^{-\frac{y^2}{2\sigma^2}}, \quad (1)$$

where $a_{\mathbf{R}_r}$ is the area inside region \mathbf{R}_r and K is the Gaussian kernel [12]. Now let $P_{\mathbf{R}_{r_{in}}, \mathbf{I}}$ denotes the image distribution corresponding to the blood within the cavity of the reference image \mathbf{I} whose delineation is given by the user (refer to the distributions depicted by the discontinuous, red curves in Fig. 2(d-f)). Furthermore, we estimate the distribution inside Γ_{out}^{nm} by $P_{\mathbf{R}_{r_{out}}^{nm}, \mathbf{I}}$ in all other frames in the sequence as shown in Fig. 2(d-f) (continuous, pink curves). Now we consider the similarity measurement, β , between image distribution $P_{\mathbf{R}_{r_{in}}, \mathbf{I}}$ corresponding to the blood in the cavity of the reference image \mathbf{I} and distribution $P_{\mathbf{R}_{r_{out}}^{nm}, \mathbf{I}}$:

$$\beta^{n,m} = \mathbf{B}(P_{\mathbf{R}_{r_{in}}, \mathbf{I}}, P_{\mathbf{R}_{r_{out}}^{nm}, \mathbf{I}}), \quad \mathbf{B}(f, g) = \int_{R^+} \sqrt{fg} dz \quad (2)$$

$\mathbf{B}(f, g)$ is the Bhattacharyya coefficient measuring the overlap (similarity) between distributions f and g . The range of the Bhattacharyya coefficient is $[0, 1]$, with 0 indicating no overlap between the distributions and 1 indicating a perfect match. The fixed $[0, 1]$ range of the Bhattacharyya coefficient affords a conveniently practical appraisal of the similarity. We expect that measurement $\beta^{n,m}$ is related to the amount of blood in the corresponding segment I_{nm} , a relationship that is demonstrated experimentally by the typical example in Fig. 2. Such similarity is reasonable because the more overlap between the image distribution within cavity and the distribution within regional segment I_{nm} , the higher the blood volume inside regional segment I_{nm} . When a regional myocardial muscle does not contract properly, the distribution of blood within $\mathbf{R}_{r_{out}}^{nm}, \mathbf{I}$, does not change and, therefore, $\beta^{n,m} \forall i \in \{1, \dots, 20\}$, can be used as a criterion to assess the myocardial function of segment I_{nm} . We then employ a linear SVM classifier and use the estimated $\beta^{n,m}$ (there are 20 $\beta^{n,m}$ for one regional segment) as features to classify regional myocardial segments as normal or abnormal.

3 LDA and Linear SVM Classifier for Regional Myocardial Abnormality Detection

We applied Linear Discriminant Analysis (LDA) to reduce the dimensionality of feature vectors, $\beta^m = \{\beta^{n,m}\}$ s.t. $n \in \{1, \dots, 20\}$, while maximizing the distance between normal and abnormal classes. This can be achieved by projecting estimated feature vectors, β^m , to a new lower-dimensional feature space of β_p^m s.t. $\beta_p^m = F_{LDA}(\beta^m)$. F_{LDA} transforms β^m to β_p^m to discriminate among abnormal and normal classes [13]. Subsequently, a linear Support Vector Machine (SVM) classifier is used to identify the decision boundary to classify the regional myocardial segments into normal and abnormal categories. The vectors near the decision boundary are called support vectors. We used linear SVM classifier to maximize the margin between the support vectors of both classes. We trained the linear SVM classifier by providing β_p^m and the associated labels of normal or abnormal obtained from ground truth by an expert radiologist.

4 Experiments

Data acquisition. A set of 2D short-axis cine magnetic resonance (MR) images of 30 subjects were acquired over the cardiac cycle on a 1.5T scanner with fast-imaging employing a steady-state acquisition (FIESTA) image sequence. The acquisition parameters were: TR=2.98 ms, TE=1.2 ms, flip angle=30 degree, and slice thickness=10 mm. The data contain 90 short-axis image sequences, each consisting of 20 functional 2D images. The results for 480 myocardial segments from apical, mid-cavity and apical were compared with ground truth classifications by an expert radiologist⁴.

Applying Linear Discriminant Analysis. After estimating image features, we applied a LDA transformation for each of regional myocardial segments individually. Fig. 3 shows the projected features β_p^m after applying LDA transformation for regional segment 3 of apical, mid-cavity and basal slices. The results show that projected features for the apical cases are more discriminative than basal and mid-cavity regions. This can be explained by the fact that there are no papillary muscles in apical slices and, therefore, estimation of the distribution corresponding to blood within cavity of apical slice is less challenging compared to basal and mid-cavity slices.

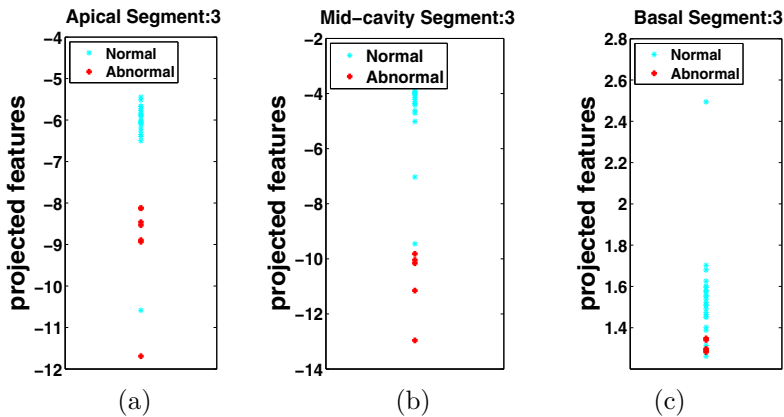


Fig. 3. projected features, β_p^m , after applying LDA transformation for regional segment 3 of apical, mid-cavity and basal slices

Linear SVM Classifier. We used 16 linear SVM classifiers to assess the 16 regional myocardial segments (normal/abnormal). Fig. 4 shows that the decision boundary separates the normal and abnormal classes using linear SVM. The decision boundary for apical is more reliable than the corresponding regional segment in the basal slice. The greater the distance between the support vectors

⁴ Among the 480 myocardial segments, 389 segments were marked as normal and 91 as abnormal.

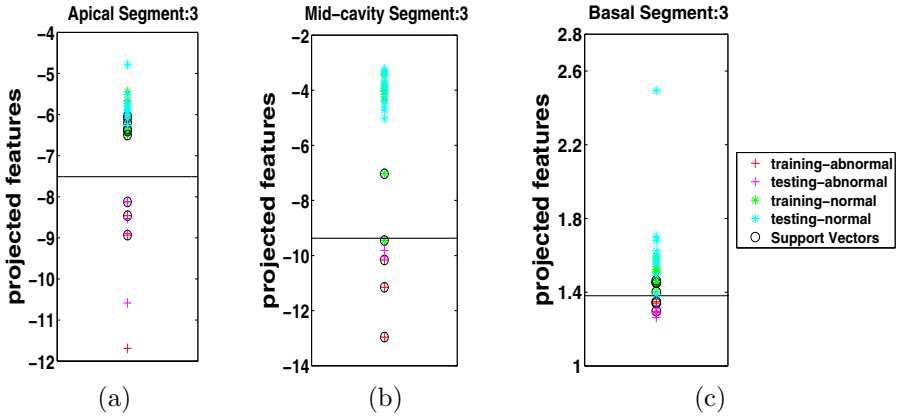


Fig. 4. Decision boundary and support vectors for regional segment 3 of apical, mid-cavity and basal slices

of normal and abnormal classes, the more reliable the decision boundary. The decision boundary in the case of mid-cavity slices suffers from misclassification because of the papillary muscles that are connected to myocardial wall.

Classification performance. We used two criteria to measure the performance of each classifier, namely the ROC , Receiver Operating Characteristics, curves with corresponding AUCs, Area Under the ROC Curve, and the Bhattacharyya measure [11] to assess the discriminative power of the features. Furthermore, we assessed the performance of the proposed approach by training our algorithm using 2/3 of the dataset and testing on the rest of the data.

ROC, AUC and Bhattacharyya measure. We show the ROC curves for classifier elements in Fig. 5. The figures show that the proposed method based on the Bhattacharyya coefficient is a reliable approach, for detecting regional abnormality in cardiac MR images. Figs. 5 (a), (b) and (c) show that apical segments are better classified than basal while basal slices are better classified than

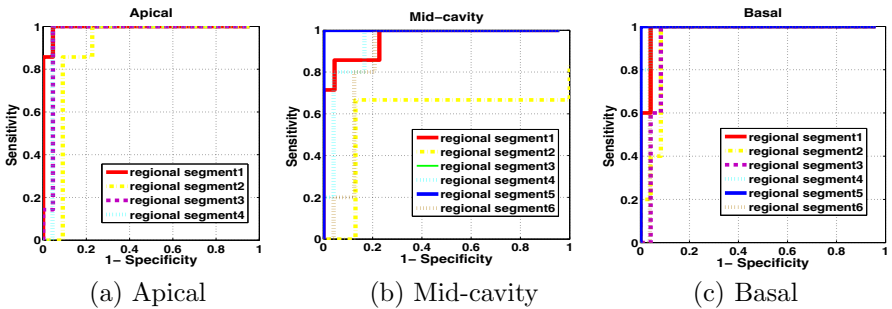


Fig. 5. Receiver operating characteristics of classifiers. The closer the curve to the left hand top corner, the better the classification performance.

Table 1. The area under the curve corresponding to Fig. 5 and the Bhattacharyya distance metric (\mathcal{B}) of normal/abnormal distributions. The higher the values, the more discriminative the ability of the classifier.

	AUC	Bhattacharyya distance metric (\mathcal{B})
Apical	0.94	0.91
Mid-cavity	0.85	0.85
Basal	0.87	0.94

mid-cavity slices. The AUCs corresponding to ROC curves in Fig. 5 are reported in Table 1. We also used the Bhattacharyya distance metric, \mathcal{B} , to evaluate the overlap between the distribution of features over normal and abnormal classes. The higher the \mathcal{B} , the more discriminative the classifier. The \mathcal{B} s in Table 1 are consistent with ROC/AUC evaluations.

Table 2. The classification accuracy computed by leaving-one-third-of-the-subjects-out. The proposed method achieved an overall classification accuracy of 91.54%.

	Sensitivity (%)	Specificity (%)	Accuracy (%)
Apex	100.0	90.91	92.86
Mid-cavity	93.33	92.93	90.48
Base	83.0	94.45	91.3

We also evaluated the performance of the classifier by computing the accuracy, specificity and sensitivity over datasets. Table 2 reports the results. The overall classification accuracy is equal to 91.5%, with a sensitivity of 92.1% and specificity of 92.8%. The highest performance was achieved for apical slices with average of 92.9% for accuracy, 100% for sensitivity, and 90.9% for specificity.

5 Conclusions

We presented a regional cardiac abnormality detection method based on the statistics of the image, which were estimated based on user-provided delineation of the first frame. Then, from this simple input, we estimated image statistics for each regional segment, and used them as features for regional heart abnormality classification. The LDA was applied to estimate projected features and a linear SVM classifier was used to classify regional LV segments into normal or abnormal classes. The experimental analysis was carried out over 90×20 segmented LV cavities of short-axis MR images obtained from 30 subjects, and demonstrated that the proposed method performs significantly better than other state-of-art methods, and can lead to a promising diagnostic support tool to assists clinicians.

References

1. Bleumink, G.S., Knetsch, A.M., Sturkenboom, M.C., Straus, S.M., Hofman, A., Deckers, J.W., Wittteman, J.C., Stricker, B.H.: Quantifying the heart failure epidemic: Prevalence, incidence rate, lifetime risk and prognosis of heart failure - the rotterdam study. *European Heart Journal* 25, 1614–1619 (2004)
2. Buckberg, G.: Left ventricular form and function: Scientific priorities and strategic planning for development of new views of disease. *Circulation* 110, e333–e336 (2004)
3. Sundar, H., Davatzikos, C., Biros, G.: Biomechanically-constrained 4D estimation of myocardial motion. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5762, pp. 257–265. Springer, Heidelberg (2009)
4. Punithakumar, K., Li, S., Ayed, I.B., Ross, I., Islam, A., Chong, J.: Heart motion abnormality detection via an information measure and bayesian filtering. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5762, pp. 373–380. Springer, Heidelberg (2009)
5. Cerqueira, M.D., Weissman, N.J., Dilsizian, V., Jacobs, A.K., Kaul, S., Laskey, W.K., Pennell, D.J., Rumberger, J.A., Ryan, T., Verani, M.: Standardized myocardial segmentation and nomenclature for tomographic imaging of the heart: A statement for healthcare professionals from the cardiac imaging committee of the council on clinical cardiology of the American Heart Association. *Circulation* 105(4), 539–542 (2002)
6. Punithakumar, K., Ayed, I.B., Islam, A., Ross, I.G., Li, S.: Regional heart motion abnormality detection via information measures and unscented kalman filtering. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6361, pp. 409–417. Springer, Heidelberg (2010)
7. Garcia-Barnes, J., Gil, D., Badiella, L., Hernández-Sabaté, A., Carreras, F., Pujadas, S., Martí, E.: A normalized framework for the design of feature spaces assessing the left ventricular function. *IEEE Transaction on Medical Imaging* 29(3), 733–745 (2010)
8. Suinesiaputra, A., Frangi, A., Kaandorp, T., Lamb, H., Bax, J., Reiber, J., Lelieveldt, B.: Automated detection of regional wall motion abnormalities based on a statistical model applied to multislice short-axis cardiac MR images. *IEEE Transaction on Medical Imaging* 28(4), 595–607 (2009)
9. Lu, Y., Radau, P., Connelly, K., Dick, A., Wright, G.: Pattern recognition of abnormal left ventricle wall motion in cardiac MR. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5762, pp. 750–758. Springer, Heidelberg (2009)
10. Mansor, S., Noble, J.: Local wall motion classification of stress echocardiography using a hidden Markov model approach. In: The IEEE International Symposium in Biomedical Imaging: From Nano to Macro, pp. 1295–1298 (2008)
11. Ben Ayed, I., Ross, I., Li, S.: Embedding overlap priors in variational left ventricle tracking. *IEEE Transaction on Medical Imaging* 28(12), 1902–1913 (2009)
12. Ben Ayed, I., Li, S., Ross, I.: A statistical overlap prior for variational image segmentation. *International Journal of Computer Vision* 85(1), 115–132 (2009)
13. Kim, H.C., Kim, D., Bang, S.Y.: Face recognition using lda mixture model. *Pattern Recogn. Lett.* 24, 2815–2821 (2003)

Identifying AD-Sensitive and Cognition-Relevant Imaging Biomarkers via Joint Classification and Regression

Hua Wang¹, Feiping Nie¹, Heng Huang¹, Shannon Risacher²,
Andrew J. Saykin², Li Shen², and ADNI*

¹ Computer Science and Engineering, University of Texas at Arlington, TX
{huawangcs, feipingnie}@gmail.com, heng@uta.edu

² Radiology and Imaging Sciences, Indiana University School of Medicine, IN
{srisache, asaykin, shenli}@iupui.edu

Abstract. Traditional neuroimaging studies in Alzheimer’s disease (AD) typically employ independent and pairwise analyses between multimodal data, which treat imaging biomarkers, cognitive measures, and disease status as isolated units. To enhance mechanistic understanding of AD, in this paper, we conduct a new study for identifying imaging biomarkers that are associated with both cognitive measures and AD. To achieve this goal, we propose a new sparse joint classification and regression method. The imaging biomarkers identified by our method are AD-sensitive and cognition-relevant and can help reveal complex relationships among brain structure, cognition and disease status. Using the imaging and cognition data from Alzheimer’s Disease Neuroimaging Initiative database, the effectiveness of the proposed method is demonstrated by clearly improved performance on predicting both cognitive scores and disease status.

1 Introduction

Neuroimaging is a powerful tool for characterizing neurodegenerative process in the progression of Alzheimer’s disease (AD). Pattern classification methods have been widely employed to predict disease status using neuroimaging measures [2,3]. Since AD is a neurodegenerative disorder characterized by progressive impairment of memory and other cognitive functions, regression models have been investigated to predict clinical scores from individual magnetic resonance imaging (MRI) and/or positron emission tomography (PET) scans [8,9]. For example, in [9], stepwise regression was performed in a pairwise fashion to relate each of MRI and FDG-PET measures of eight candidate regions to each of four Rey’s Auditory Verbal Learning Test (RAVLT) memory scores.

* Data collection and sharing for this project was funded by the Alzheimer’s Disease Neuroimaging Initiative (ADNI) (U01 AG024904,adni.loni.ucla.edu). HW and FN contributed equally to this work. HW and HH were supported by NSF-CNS 0923494, NSF-IIS 1041637, NSF-CNS 1035913. SR, AS and LS were supported in part by NIBIB R03 EB008674, NIA 1RC 2AG036535, CTSI-IUSM/CTR (RR025761), NIA P30 AG10133, and NIA R01 AG19771.

Predicting disease status and predicting memory performance, using neuroimaging data, are both important learning tasks. Prior research typically studied these tasks *separately*. One example is to first determine disease-relevant cognitive scores and then identify imaging biomarkers associated with these scores so that interesting pathways from brain structure to cognition to symptom can potentially be discovered. However, a specific cognitive function could be related to multiple imaging measures associated with different biological pathways (some of them are not related to AD). As a result, the identified imaging biomarkers are not necessarily all disease specific. To have a better understanding of the underlying mechanism specific to AD, an interesting topic would be to only discover imaging biomarkers associated with both cognitive function and AD status.

To identify AD-sensitive and cognition-relevant imaging biomarkers, we propose a new joint classification and regression learning model to simultaneously performing two heterogeneous tasks, *i.e.*, imaging-to-disease classification and imaging-to-cognition regression. We use magnetic resonance imaging (MRI) measures as predictors and cognitive memory scores and disease status as response variables. For each individual regression or classification task, we employ a multi-task learning model [1] in which tasks for predicting different memory performances (or those for predicting AD and control dummy variables in classification) are considered as homogeneous tasks. Different to LASSO and other related methods that mainly find the imaging features correlated to each individual memory score, our method selects the imaging features that tend to play an important role on influencing multiple homogenous tasks.

Our new method utilizes the sparse regularization to perform imaging biomarker selection and learn a sparse parameter matrix under a unified framework that integrates both heterogeneous and homogenous tasks. Specifically, by recognizing that the formation [6] and maintenance [4] of memory are synergically accomplished by a few brain areas, such as medial temporal lobe structures, medial and lateral parietal, as well as prefrontal cortical areas, we use the $\ell_{2,1}$ -norm regularization to select features that can predict most memory scores and classify AD versus control. Empirical comparison with the existing methods demonstrates that the proposed method not only yields improved performance on predicting both cognitive scores and disease status, but also discovers a small set of AD-sensitive and cognition-relevant biomarkers in accordance with prior findings.

2 Sparse Model for Joint Classification and Regression

When we study either regression or classification via a multi-task learning model, given a set of input variables, (*i.e.*, features, such as imaging biomarkers), we are interested in learning a set of related models (*e.g.*, associations between image biomarkers and cognitive scores) for predicting multiple homogenous tasks (such as predicting cognitive scores). Since these homogenous tasks are typically interrelated, they share a common input space. As a result, it is desirable to learn all the models jointly rather than treating each task as an independent one.

Such multi-task learning methods can help discover robust patterns, especially when significant patterns in a single task become outliers for other tasks, and potentially increase the predictive power.

To identify AD-sensitive and cognition-relevant biomarkers from imaging data, we formulate a new problem to jointly learn two heterogeneous tasks: classification and regression. We propose a new sparse model for joint classification and regression to perform multivariate regression for cognitive memory scores predictions and logistic regression for disease classification tasks simultaneously.

Notation. We write matrices and vectors as bold uppercase and lowercase letters respectively. Given a matrix $\mathbf{M} = [m_{ij}]$, we denote its i -th row as \mathbf{m}^i and j -th column as \mathbf{m}_j . The Frobenius norm of the matrix \mathbf{M} is denoted as $\|\mathbf{M}\|_{\text{F}}$, and the $\ell_{2,1}$ -norm [5] of \mathbf{M} is defined as $\|\mathbf{M}\|_{2,1} = \sum_i \sqrt{\sum_j m_{ij}^2} = \sum_i \|\mathbf{m}^i\|_2$.

2.1 Objective of Sparse Joint Classification and Regression

First, logistic regression is used for disease classification. Given the training data $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n] \in \mathbb{R}^{d \times n}$, each data point \mathbf{x}_i is associated with a label vector $\mathbf{y}^i = [y_{i1}, \dots, y_{ic_1}] \in \mathbb{R}^{c_1}$. If \mathbf{x}_i belongs to the k -th class, $y_{ik} = 1$, otherwise $y_{ik} = 0$. We write $\mathbf{Y} = [(\mathbf{y}^1)^T, \dots, (\mathbf{y}^n)^T]^T \in \mathbb{R}^{n \times c_1}$. In traditional multi-class logistic regression, under a projection matrix $\mathbf{W} \in \mathbb{R}^{d \times c_1}$, we have

$$p(k | \mathbf{x}_i, \mathbf{W}) = \frac{e^{\mathbf{w}_k^T \mathbf{x}_i}}{\sum_{l=1}^{c_1} e^{\mathbf{w}_l^T \mathbf{x}_i}} \implies p(\mathbf{y}^i | \mathbf{x}_i, \mathbf{W}) = \prod_{k=1}^{c_1} \left(\frac{e^{\mathbf{w}_k^T \mathbf{x}_i}}{\sum_{l=1}^{c_1} e^{\mathbf{w}_l^T \mathbf{x}_i}} \right)^{y_{ik}},$$

where $p(k | \mathbf{x}_i, \mathbf{W})$ is the probability that \mathbf{x}_i belongs to the k -th class, and $p(\mathbf{y}^i | \mathbf{x}_i, \mathbf{W})$ is the probability that \mathbf{x}_i is associated with the given label \mathbf{y}^i . Therefore, the multi-class logistic loss that maximizes the Log-likelihood can be achieved by minimizing:

$$l_1(\mathbf{W}) = -\log \prod_{i=1}^n p(\mathbf{y}^i | \mathbf{x}_i, \mathbf{W}) = \sum_{i=1}^n \sum_{k=1}^{c_1} \left(y_{ik} \log \sum_{l=1}^{c_1} e^{\mathbf{w}_l^T \mathbf{x}_i} - y_{ik} \mathbf{w}_k^T \mathbf{x}_i \right). \quad (1)$$

In AD classification, we have two classes, *i.e.*, AD and health control (HC).

Second, we use multivariate least square regression to predict cognitive scores, which minimizes:

$$l_2(\mathbf{P}) = \|\mathbf{X}^T \mathbf{P} - \mathbf{Z}\|_{\text{F}}^2, \quad (2)$$

where \mathbf{X} is the data matrix, $\mathbf{Z} = [(\mathbf{z}^1)^T, \dots, (\mathbf{z}^n)^T]^T \in \mathbb{R}^{n \times c_2}$ is the label matrix for the c_2 regression tasks, and $\mathbf{P} \in \mathbb{R}^{d \times c_2}$ is the projection matrix.

The objective for joint classification and regression to identify AD-sensitive and cognition-relevant imaging biomarkers can now be formulated as follows:

$$\min J(\mathbf{V}) = l_1(\mathbf{W}) + l_2(\mathbf{P}) + \gamma \|\mathbf{V}\|_{2,1}, \quad (3)$$

where $\mathbf{V} = [\mathbf{W} \ \mathbf{P}] \in \mathbb{R}^{d \times (c_1 + c_2)}$. Thanks to the $\ell_{2,1}$ -norm regularization on \mathbf{V} [1], the biomarkers are identified across all tasks so that they are not only correlated to cognitive scores but also discriminative to disease status.

2.2 An Efficient Iterative Algorithm

Due to the non-smoothness of the $\ell_{2,1}$ -norm term, J in Eq. (3) is hard to solve in general. Thus we derive an efficient iterative algorithm as follows.

Taking the derivatives of J w.r.t. \mathbf{W} and \mathbf{P} , we set them to be zeros:

$$\frac{\partial J}{\partial \mathbf{W}} = \frac{\partial l_1(\mathbf{W})}{\partial \mathbf{W}} + 2\gamma \mathbf{D}\mathbf{W} = 0, \quad \frac{\partial J}{\partial \mathbf{P}} = 2\mathbf{X}\mathbf{X}^T\mathbf{P} - 2\mathbf{X}\mathbf{Z} + 2\gamma\mathbf{D}\mathbf{P} = 0, \quad (4)$$

where \mathbf{D} is a diagonal matrix whose k -th diagonal element is $\frac{1}{2\|\mathbf{v}^k\|_2}$. Because \mathbf{D} depends on \mathbf{V} , it is also an unknown variable. Following standard optimization procedures in statistical learning, we alternately optimize \mathbf{V} and \mathbf{D} .

Algorithm 1. An efficient algorithm to solve Eq. (3)

Input: $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n] \in \mathbb{R}^{d \times n}$, $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_n]^T \in \mathbb{R}^{n \times c_1}$, and $\mathbf{Z} = [\mathbf{z}_1, \dots, \mathbf{z}_n]^T \in \mathbb{R}^{n \times c_2}$.

1. Initialize $\mathbf{W} \in \mathbb{R}^{d \times c_1}$, $\mathbf{P} \in \mathbb{R}^{d \times c_2}$, and let $\mathbf{V} = [\mathbf{W} \mathbf{P}] \in \mathbb{R}^{d \times (c_1+c_2)}$;

while not converge **do**

2. Calculate the diagonal matrix \mathbf{D} , of which the k -th element is $\frac{1}{2\|\mathbf{v}^k\|_2}$;

3. Update \mathbf{w} by $\mathbf{w} - \mathbf{B}^{-1}\mathbf{a}$, where $(d \times (p-1) + u)$ -th element of $\mathbf{a} \in \mathbb{R}^{dc_1 \times 1}$ is

$\frac{\partial [l_1(\mathbf{W}) + \gamma \text{tr}(\mathbf{W}^T \mathbf{D}\mathbf{W})]}{\partial \mathbf{W}_{up}}$ for $1 \leq u \leq d, 1 \leq p \leq c_1$, the $(d \times (p-1) + u, d \times (q-1) + v)$ -th

element of $\mathbf{B} \in \mathbb{R}^{dc_1 \times dc_1}$ is $\frac{\partial [l_1(\mathbf{W}) + \gamma \text{tr}(\mathbf{W}^T \mathbf{D}\mathbf{W})]}{\partial \mathbf{W}_{up} \partial \mathbf{W}_{vq}}$ for $1 \leq u, v \leq d$ and $1 \leq p, q \leq c_1$.

Construct the updated $\mathbf{W} \in \mathbb{R}^{d \times c_1}$ by the updated vector $\mathbf{w} \in \mathbb{R}^{dc_1}$, where the (u, p) -th element of \mathbf{W} is the $(d \times (p-1) + u)$ -th element of \mathbf{w} ;

4. Update \mathbf{P} by $\mathbf{P} = (\mathbf{X}\mathbf{X}^T + \gamma\mathbf{D})^{-1}\mathbf{X}\mathbf{Z}$;

5. Update \mathbf{V} by $\mathbf{V} = [\mathbf{W} \mathbf{P}]$;

end

Output: $\mathbf{W} \in \mathbb{R}^{d \times c_1}$ and $\mathbf{P} \in \mathbb{R}^{d \times c_2}$.

First, we randomly initialize $\mathbf{V} \in \mathbb{R}^{d \times (c_1+c_2)}$, upon which we calculate \mathbf{D} . After obtaining \mathbf{D} , we update the solution $\mathbf{V} = [\mathbf{W} \mathbf{P}]$ using Eq. (4). To be more precise, \mathbf{P} is updated by $\mathbf{P} = (\mathbf{X}\mathbf{X}^T + \gamma\mathbf{D})^{-1}\mathbf{X}\mathbf{Z}$. Because we cannot update \mathbf{W} with a closed form solution upon Eq. (4), we employ Newton's method to obtain updated \mathbf{W} by solving the following problem: $\min_{\mathbf{W}} l_1(\mathbf{W}) + \gamma \text{tr}(\mathbf{W}^T \mathbf{D}\mathbf{W})$.

Once we obtain the updated $\mathbf{V} = [\mathbf{W} \mathbf{P}]$, we can calculate \mathbf{D} . This procedure repeats until convergence. The detailed algorithm is summarized in Algorithm 1, whose convergence is proved as following.

Lemma 1. For any vector \mathbf{v} and \mathbf{v}_0 , we have $\|\mathbf{v}\|_2 - \frac{\|\mathbf{v}\|_2^2}{2\|\mathbf{v}_0\|_2} \leq \|\mathbf{v}_0\|_2 - \frac{\|\mathbf{v}_0\|_2^2}{2\|\mathbf{v}_0\|_2}$. Proof is available in [5].

Theorem 1. Algorithm 1 decreases the objective value of J in every iteration.

Proof. In each iteration, denote the updated \mathbf{W} as $\widetilde{\mathbf{W}}$, the updated \mathbf{P} as $\widetilde{\mathbf{P}}$, thus the updated \mathbf{V} is $\widetilde{\mathbf{V}} = [\widetilde{\mathbf{W}} \widetilde{\mathbf{P}}]$. According to step 3 of Algorithm 1, we have

$$l_1(\widetilde{\mathbf{W}}) + \gamma \text{tr}(\widetilde{\mathbf{W}}^T \mathbf{D}\widetilde{\mathbf{W}}) \leq l_1(\mathbf{W}) + \gamma \text{tr}(\mathbf{W}^T \mathbf{D}\mathbf{W}). \quad (5)$$

According to step 4 we know that

$$l_2(\tilde{\mathbf{P}}) + \gamma \operatorname{tr}(\tilde{\mathbf{P}}^T \mathbf{D} \tilde{\mathbf{P}}) \leq l_2(\mathbf{P}) + \gamma \operatorname{tr}(\mathbf{P}^T \mathbf{D} \mathbf{P}). \quad (6)$$

According to the definition of \mathbf{D} and Lemma 1, we have the following inequality:

$$\begin{aligned} \sum_{k=1}^d \|\tilde{\mathbf{v}}^k\|_2 - \sum_{k=1}^d \frac{\|\tilde{\mathbf{v}}^k\|_2^2}{2\|\mathbf{v}^k\|_2} &\leq \sum_{k=1}^d \|\mathbf{v}^k\|_2 - \sum_{k=1}^d \frac{\|\mathbf{v}^k\|_2^2}{2\|\mathbf{v}^k\|_2} \\ \Rightarrow \gamma \sum_{k=1}^d \|\tilde{\mathbf{v}}^k\|_2 - \gamma \operatorname{tr}(\tilde{\mathbf{V}}^T \mathbf{D} \tilde{\mathbf{V}}) &\leq \gamma \sum_{k=1}^d \|\mathbf{v}^k\|_2 - \gamma \operatorname{tr}(\mathbf{V}^T \mathbf{D} \mathbf{V}). \end{aligned} \quad (7)$$

Because $\operatorname{tr}(\mathbf{V}^T \mathbf{D} \mathbf{V}) = \operatorname{tr}(\mathbf{W}^T \mathbf{D} \mathbf{W}) + \operatorname{tr}(\mathbf{P}^T \mathbf{D} \mathbf{P})$, by adding Eqs. (5-7) at the both sides, we arrive at

$$l_1(\tilde{\mathbf{W}}) + l_2(\tilde{\mathbf{P}}) + \gamma \sum_{k=1}^d \|\tilde{\mathbf{v}}^k\|_2 \leq l_1(\mathbf{W}) + l_2(\mathbf{P}) + \gamma \sum_{k=1}^d \|\mathbf{v}^k\|_2 \quad (8)$$

Thus, Algorithm 1 decreases the value of J in Eq. (3) in every iteration. \square

Because J in Eq. (3) is obviously lower-bounded by 0, Theorem 1 guarantees the convergence of Algorithm 1. In addition, because J is convex, Algorithm 1 converges at the global optimum of the problem.

3 Experimental Results

We evaluate our method by applying it to the Alzheimer’s Disease Neuroimaging Initiative (ADNI) cohort. The goal is to select a compact set of AD-sensitive and cognition-relevant imaging biomarkers while maintaining high predictive power.

Data preparation. We downloaded data from the ADNI database (<http://adni.loni.ucla.edu>). We used baseline MRI data, from which we extracted 56 volumetric and cortical thickness values (Fig. 1) using FreeSurfer (<http://surfer.nmr.mgh.harvard.edu>), as described in [7]. We included memory scores from three different cognitive assessments including Mini-Mental State Exam (MMSE), Rey’s Auditory Verbal Learning Test (RAVLT), and TRAILS. Details about these assessments are available in the ADNI procedure manuals (<http://www.adni-info.org/Scientists/ProceduresManuals.aspx>).

3.1 Biomarker Identification

The proposed method aims to identify imaging biomarkers that are associated with both disease status and cognitive scores in a joint classification and regression framework. Here we first examine the identified biomarkers. Fig. 1 shows a summarization of selected features for the three experiments (one for each type of cognitive scores) where the regression/classification weights are color-mapped

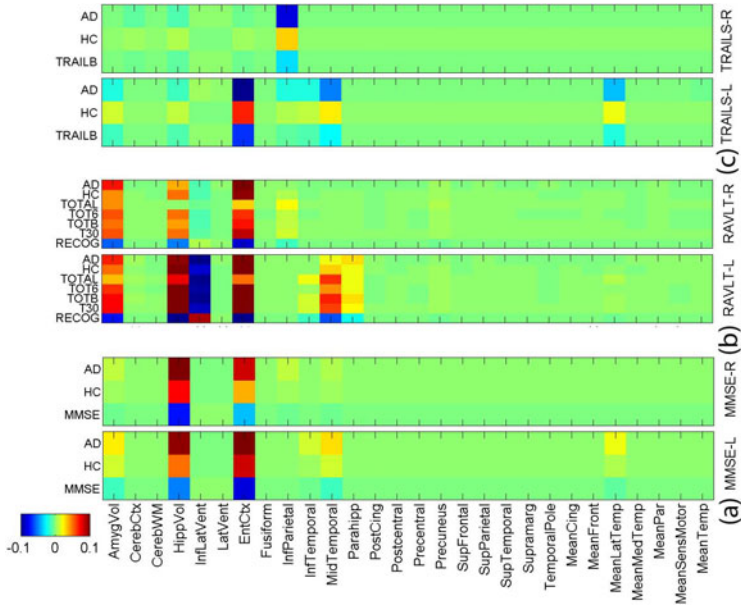


Fig. 1. Weight maps of the joint classification and regression tasks. One binary classification task for AD and HC. Three different groups of cognitive scores for regression: (a) MMSE score, (b) RAVLT score, (c) TRAILS score. “-L” indicates the FreeSurfer biomarkers at the left side, and “-R” indicates those at the right side.

for each feature and each task. Fig. 2 visualizes the cortical maps of selected features for both classification and regression in different tasks.

Fig. 1 and Fig. 2 show that a small set of MRI measures are identified, including hippocampal volume (HippVol), entorhinal cortex thickness (EntCtx), amygdala volume (AmygVol), inferior parietal gyrus thickness (InfParietal), and middle temporal gyrus thickness (MidTemporal). These are all well-known AD-relevant biomarkers. Our method also shows that these markers are jointly associated with one or more memory scores. Although we know that MRI measures, cognitive scores and diagnosis are highly correlated, the complex relationships among them remain to be discovered for a better understanding of AD mechanism. This is one major focus of our work. As shown in Fig. 1, different AD-sensitive MRI measures could be related to different cognitive tasks. The proposed sparse method for joint classification and regression enables us to sort out MRI-cognition relationships while focusing on AD-sensitive markers.

3.2 Improved Prediction Performance

Now we evaluate the performance of joint classification and regression for AD detection and cognitive score prediction using MRI data. We performed standard 5-fold cross-validation, where the parameter γ of our method in Eq. (3) was fine

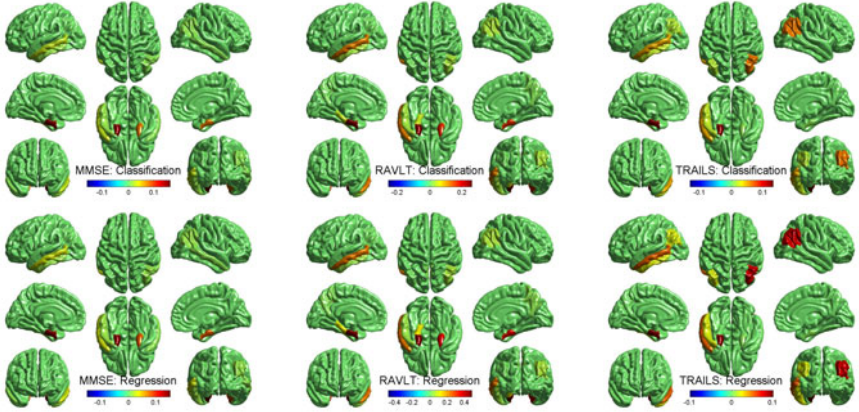


Fig. 2. Cortical map of selected features for cognitive score prediction using FreeSurfer measures in the three joint classification and regression tasks

tuned in the range of $\{10^{-5}, \dots, 1, \dots, 10^5\}$ by an internal 5-fold cross-validation in the training data of each of the 5 trials. For classification, we compared the proposed method against two baseline methods including logistic regression and support vector machine (SVM). For SVM, we implemented three different kernels including linear, polynomial and Gaussian kernels. For polynomial kernel, we searched the best results when the polynomial order varied in the range of $\{1, 2, \dots, 10\}$; for Gaussian kernel, we fine tuned the parameter α in the same range as that for our method and fixed parameter C as 1. For regression, we compared our method against two widely used methods including multivariate regression and ridge regression. For the latter, we fine tuned its parameter in the same range as that for our method. The results are reported in Table [1](#).

Table [1](#) shows that our method performs clearly better than both logistic regression and SVM, which are consistent with our motivations in that our method classifies participants using the information from not only MRI measures but also the reinforcement by cognitive score regression. In addition, the cognitive score regression performances of our method measured by root mean squared error (RMSE) outperform both multivariate regression and ridge regression, supporting the usefulness of joint classification and regression from another perspective. Ridge regression achieves close but slightly worse regression performance. However, it lacks the ability to identify relevant imaging markers. All these observations demonstrate the effectiveness of the proposed method in improving the performances of both AD detection and cognitive score prediction.

Mild cognitive impairment (MCI) is thought to be the prodromal stage of AD. Including MCI in this type of analyses will be an interesting future direction to help biomarker discovery for early detection of AD. We performed an initial analysis on three-class classification for AD, MCI and HC: the accuracy of our method was 0.663 and the best of other tested methods was 0.615. Apparently this is a much harder task and warrants further thorough investigation.

Table 1. Comparison of classification and regression performance

Memory score	# subjects	# AD	# HC	Our method		Classification accuracy		RMSE (mean \pm std)	
				Classification accuracy	Regression RMSE	Logistic regression	SVM	Multivariate regression	Ridge regression
MMSE	378	175	203	0.881	0.034 \pm 0.002	0.832	0.783 (linear kernel)	0.041 \pm 0.003	0.039 \pm 0.004
RAVLT	371	172	199	0.884	0.019 \pm 0.001		0.839 (Polynomial kernel)	0.028 \pm 0.002	0.024 \pm 0.003
TRAILS	369	166	203	0.864	0.043 \pm 0.002		0.796 (Gaussssian kernel)	0.049 \pm 0.003	0.046 \pm 0.003

4 Conclusions

We have proposed a new sparse regression model for joint classification and regression and applied it to the ADNI cohort for identifying AD-sensitive and cognition-relevant imaging biomarkers. Our methodological contributions are threefold: 1) proposing a new learning model, joint classification and regression learning, to identify disease-sensitive and task-relevant biomarkers for analyzing multimodal data; 2) employing structural sparsity regularization to integrate heterogeneous and homogenous tasks in a unified multi-task learning framework; 3) deriving a new efficient optimization algorithm to solve our non-smooth objective function, and coupling this with rigorous theoretical analysis on global optimum convergency. Empirical comparison with the existing methods demonstrates that our method not only yields improved performance on predicting both cognitive scores and disease status using MRI data, but also discovers a small set of AD-sensitive and cognition-relevant imaging biomarkers in accordance with prior findings.

References

- Argyriou, A., Evgeniou, T., Pontil, M.: Multi-task feature learning. In: NIPS 2007, pp. 41–48 (2007)
- Fan, Y., Batmanghelich, N., Clark, C.M., Davatzikos, C.: Spatial patterns of brain atrophy in MCI patients, identified via high-dimensional pattern classification, predict subsequent cognitive decline. *Neuroimage* 39(4), 1731–1743 (2008)
- Hinrichs, C., Singh, V., Mukherjee, L., Xu, G., Chung, M., Johnson, S.: Spatially augmented LPboosting for AD classification with evaluations on the ADNI dataset. *Neuroimage* 48(1), 138–149 (2009)
- Moscovitch, M., Nadel, L., Winocur, G., Gilboa, A., Rosenbaum, R.: The cognitive neuroscience of remote episodic, semantic and spatial memory. *Curr. Opin. Neurobiol.* 16(2), 179–190 (2006)
- Nie, F., Huang, H., Cai, X., Ding, C.: Efficient and robust feature selection via joint $l_{2,1}$ -norms minimization. In: NIPS 2010, pp. 1813–1821 (2010)
- Scoville, W., Milner, B.: Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery & Psychiatry* 20(1), 11 (1957)
- Shen, L., Kim, S., Risacher, S.L., Nho, K., Swaminathan, S., West, J.D., Foroud, T., Pankratz, N., Moore, J.H., Sloan, C.D., Huentelman, M.J., Craig, D.W., DeChairo, B.M., Potkin, S.G., Jack, C.R., Weiner, M.W., Saykin, A.J., ADNI: Whole genome association study of brain-wide imaging phenotypes for identifying quantitative trait loci in MCI and AD: A study of the ADNI cohort. *Neuroimage* 53(3), 1051–1063 (2010)

8. Stonnington, C.M., Chu, C., Kloppel, S., Jack Jr., C.R., Ashburner, J., Frackowiak, R.S.: Predicting clinical scores from magnetic resonance scans in alzheimer's disease. *Neuroimage* 51(4), 1405–1413 (2010)
9. Walhovd, K., Fjell, A., Dale, A., McEvoy, L., Brewer, J., Karow, D., Salmon, D., Fennema-Notestine, C.: Multi-modal imaging predicts memory performance in normal aging and cognitive decline. *Neurobiol Aging* 31(7), 1107–1121 (2010)

Biological Indexes Based Reflectional Asymmetry for Classifying Cutaneous Lesions

Zhao Liu¹, Lyndon Smith¹, Jiulai Sun¹, Melvyn Smith¹, and Robert Warr²

¹ Machine Vision Lab, University of the West of England, Bristol, UK

² Plastic Surgery, Frenchay Hospital, NHS, Bristol, UK

Abstract. This paper proposes a novel reflectional asymmetry descriptor to quantize the asymmetry of the cutaneous lesions for the discrimination of malignant melanoma from benign nevi. A pigmentation elevation model of the biological indexes is first constructed, and then the asymmetry descriptor is computed by minimizing the histogram difference of the global point signatures of the pigmentation model. Melanin and Erythema Indexes are used instead of the original intensities in colour space to characterize the pigmentation distribution of the cutaneous lesions. 311 dermoscopy images are used to validate the algorithm performance, where 88.50% sensitivity and 81.92% specificity have been achieved when employing an SVM classifier.

1 Introduction

An abnormal reproduction of biological cells within body organs and tissues usually forms an asymmetric shape or appearance, with a potential to become metastatic and aggressive spreading thorough the body. Malignant melanoma, which accounts for 75% mortality caused by skin cancers [1], is one of these examples. Asymmetry therefore has been demonstrated as one of the most important features to quantify the shape and structure of the lesion. Stoecker et al. [2] determined the reflectional asymmetry of the lesion by area differences across the principal axes computed from the moment of inertia. Similarly, Seidenari et al. [3] evaluated the lesion asymmetry by calculating the area differences with respect to 128 axes through the lesion centre. Stanganelli et al. [4] applied the size function to separately quantize the asymmetry of the lesion in terms of boundary, shape and colour to achieve a better classification results.

Most of the above approaches have shown the discrimination power of asymmetry in melanoma identification. However, they share some similar shortcomings. Firstly, most of the existing asymmetry descriptors only evaluate the extrinsic shape, but ignore the asymmetry of the inhomogeneous pigmentation inside the lesion. Secondly, since asymmetry measure and symmetry axis are normally defined separately, the final results might not be optimized. Thirdly, clinically acquired lesion images may contain complex distortions due to the factors like sensor positions and lighting conditions. Normal extrinsic asymmetry descriptors greatly vary with them.

This paper proposed a new intrinsic reflectional asymmetry descriptor to simultaneously quantize the shape and the pigmentation distributions of the cutaneous

lesions with robustness to various deformations. We compute the melanin index and erythema index from an RGB image, and then map these biology values to the Z-axis to build a pigmentation elevation model. The reflectional asymmetry of the lesion is defined by minimizing the histogram difference of the global point signatures (GPSs) of the pigmentation model. Compared with those shape dependent asymmetry detectors, the proposed asymmetry descriptor is more efficient in describing the abnormality of the lesion as it simultaneously integrates the shape and pigmentation information. Moreover, melanin and erythema indexes reveal the pathological tissue conditions, which proved more useful in characterizing the irregularity of the pigmentation of the cutaneous lesions. We have verified that the proposed GPSs-based asymmetry descriptor is invariant to 2D rigid transformations. It is also robust to the non-rigid deformations, which is ideal for characterizing the intrinsic asymmetry of the skin lesions.

2 Biology Pigmentation of the Skin Lesions

Melanin and haemoglobin are two primary chromophore components determining the colour of human skin. Excessive ultraviolet radiation from the sun may cause irregular melanin growth in horizontal and vertical directions. When the aggressive melanocytes reach the vascular system, they become malignant and are easily broadcast to the whole body. Therefore the irregularity of both melanin and haemoglobin are internal factors for skin cancer diagnosis, though they are normally presented as colour values acquired by optical cameras.

Melanin Index (MI) and Erythema Index (EI), reflecting the pigmentation and vascular blood status of human skin, can be approximated from Red and Green channels of an RGB image. Regarding the skin reflectance is in Green (~560 nm) (high hemoglobin absorption) and Red (~650nm) (low hemoglobin absorption) spectral ranges, the approximations of MI and EI can be defined [5] as,

$$MI = 100 * \log_{10}(1/R) \quad (1)$$

$$EI = 100 * (\log_{10}(1/G) - 1.44 * \log_{10}(1/R)) \quad (2)$$

where R and G are intensities in red and green channels respectively. Fig.1 shows the MI and EI mappings of a malignant melanoma (MM) and a benign nevus (BN).

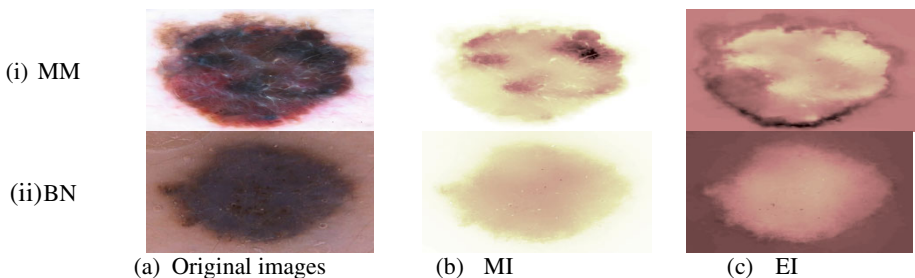


Fig. 1. Melanin and Erythema mappings of (a) MM, (b) BN

3 Asymmetry Descriptor

3.1 Pigmentation Elevation Model

A few pre-processing steps are applied to remove the hairs on the skin [6], flatten the homogeneous regions, and isolate the lesions from the surrounding normal skin [7]. Then MI and EI are calculated to build the pigmentation elevation model by mapping MI/EI to Z-axis, and the shape information of the isolated lesion as X and Y axes.

In order to enable the shape and pigmentation giving the comparable contributions, the pigmentation model is scaled down in a bound box. Suppose X_{lesion} has larger range than Y_{lesion} , then they are first normalized as $X'_{\text{lesion}} = X_{\text{lesion}} / (\max(X_{\text{lesion}}) - \min(X_{\text{lesion}}))$ and $Y'_{\text{lesion}} = (r * Y_{\text{lesion}}) / (\max(Y_{\text{lesion}}) - \min(Y_{\text{lesion}}))$. As a result, the range of X'_{lesion} is 1, and that of Y'_{lesion} is r , where $r = Y_{\text{lesion}} / X_{\text{lesion}}$. Next normalize Z-axis to the range of [0 1]. Because the maximum and the minimum values in MI and EI vary among images, conventional normalization yields non-uniform results. So we calculate the range of MI and EI indexes according to (1) and (2), where the range of MI is [-240.82, 0] and that of EI is [-240.82, 346.78]. Then the normalized EI and MI images can be calculated below to ensure the effectiveness of the pigmentation normalization is equivalent to every single image,

$$MI_{\text{norm}} = \frac{MI - (-240.82)}{240.82} = 0.0042MI + 1 \quad (3)$$

$$EI_{\text{norm}} = \frac{EI - (-240.82)}{587.6} = 0.0017EI + 0.4098 \quad (4)$$

Fig.2 shows the pigmentation elevation models calculated from MI and EI in Fig.1. The outer boundary of the model stands for the shape asymmetry of the skin lesion, and the pigmentation asymmetry of the lesion can be reflected by the distribution of the pigmentation model along the Z-direction.

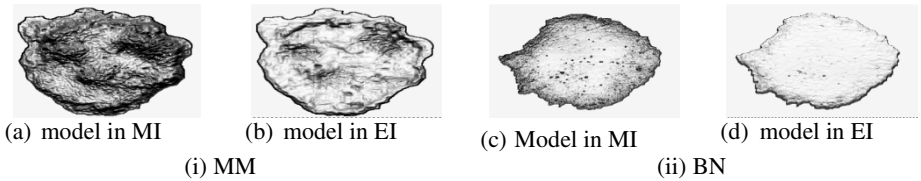


Fig. 2. Pigmentation elevation model from lesion images in Fig.1. (i) MM, (ii) BN

3.2 Reflectional Asymmetry in Histograms

The proposed GPSs-based asymmetry descriptor is derived from the Laplace-Beltrami operator, which is defined as the divergence of the gradient on a surface function in Riemannian manifold M . The Laplace-Beltrami operator L can be approximated by the graph Laplacian matrix using the heat kernel [8], where the eigenvectors of the Laplacian matrix embed the points on M into lower dimensional representations. Since the Laplace-Beltrami operator is symmetric negative and semidefinite, it has an eigen-decomposition $L\phi_i = \lambda_i\phi_i$ ($0 = \lambda_0 < \lambda_1 < \dots < \lambda_i$), where ϕ_i are eigenfunctions and λ_i are the corresponding eigenvalues. The GPSs can then be calculated [9] as:

$$gps = \left(\frac{\phi_1}{\sqrt{\lambda_1}}, \frac{\phi_2}{\sqrt{\lambda_2}}, \frac{\phi_3}{\sqrt{\lambda_3}}, \dots, \frac{\phi_i}{\sqrt{\lambda_i}}, \dots \right) \quad (5)$$

GPSs are used because they can integrate shape and colour asymmetry detections on the pigmentation model simultaneously. Moreover, it has been proved that GPSs reflect the intrinsic metric property of the object and are robust to the metric deformations [9] which might be introduced by the camera positions or various lighting conditions during the data acquisition.

In our work, M is the pigmentation model in 3-dimensional space. The eigenvectors corresponding to the first few eigenvalues determine the optimal embeddings. Since the first eigenvector with $\lambda_0 = 0$ generates a constant function and the eigenvectors associated with repeated eigenvalues are not stable for small non-isometric perturbations [9], we restrict our searching for reflectional asymmetry detection in the first 6 eigenfunctions with non-zero and non-repeated eigenvalues.

GPSs uniquely determine the metric of the manifold. So M is intrinsically symmetric if there is an associated self-mapping $T: M \rightarrow M$, making both ϕ_i and $\phi_i \circ T$ the eigenvectors of L . Suppose gps_i is the i^{th} ($1 \leq i \leq 6$) component in the GPSs, and $T = \{t_1, t_2, \dots, t_6\}$ is the self-mapping. For a complete symmetry object, the GPSs with non-repeated eigenvalues only holds two possibilities along the reflectional symmetry axis as $gps_i \circ T_i = gps_i$ and $gps_i \circ T_i = -gps_i$. Thus T can be determined by a sign sequence with either positive (+1) or negative (-1). Skin lesions are imperfect or non-symmetric objects, so the complete symmetry measure $|gps \circ T| = |gps|$ could not be fulfilled. Thus we generate a region-based reflectional asymmetry descriptor in histogram to quantify the asymmetry of the skin lesions in the GPS spaces.

We first defined the gravity centre of a lesion, and then the entire lesion is segmented into 180 segments around the polar coordinate across the centre. For each segment, we built the histogram of GPS with 100 bins, thus the descriptor of each segment l in signature i can be represented as (6).

$$Des_{i,l}(T) = \frac{1}{N_l} \sum_{n=1}^{100} f(gps_{i,l}^n(T)) * v(gps_{i,l}^n(T)) \quad l = 1, 2, \dots, 180 \quad (6)$$

where f represents the frequency counts in each bin, v is the bin location of $gps_{i,l}(T)$ in histogram and N_l is the number of pixels in each segment. For every $gps_i(T)$, the $Des_i(T)$ value can be plotted from 0 to π . As the principal axis must exist in these 180 segments, we assume one segment as the principal axis at each time, and translate part of the $Des_i(T)$ to ensure 90 elements on both sides along the axis. Considering there are six GPSs, the asymmetric degree of a lesion can be quantified by minimizing the Euclidean distance between the left-right sides of the histogram,

$$Asy(T) = \min \left(\sum_{i=1}^6 \sum_{l=1}^{90} \|Des_{i,l}^L(T) - Des_{i,l}^R(T)\|_2^2 \right) \quad (7)$$

where $Des_{i,l}^L$ represents the left side and $Des_{i,l}^R$ stands for the right part of the histogram. Because the asymmetry descriptor is a function of the sign sequence T , the minimum asymmetry measure also indicates the potential optimal reflectional axis. Moreover, since the GPSs integrate the shape and pigmentation asymmetry detections simultaneously, it avoids yielding two different symmetry planes when the asymmetric appearance of shape and colour are analyzed separately.

Fig.3(i) shows the first 6 GPSs with the optimal sign sequence T from the MI images of the lesions in Fig.1. The BN shows approximate symmetric appearance in a

given direction for all $gps(T)$, whereas it is difficult to find an appropriate symmetry plane fit for the six $gps(T)$ of the MM. The first two $gps(T)$ indicate shape asymmetry, while higher orders of $gps(T)$ reflect the pigmentation distribution inside the lesion. Fig.3(ii) plots the optimal translated histograms with the minimum asymmetry measure. The BN shows symmetry on left-right side of the histograms, while the optimal translated histograms of the MM greatly fluctuate. It results in a large asymmetry measure which can be used for the classification purpose.

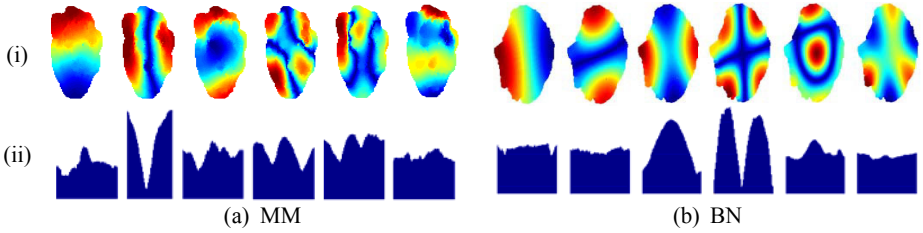


Fig. 3. Asymmetry descriptors of the MM and the BN in Fig.1. (i) The first 6 GPSs with the optimal sign sequence T. (ii) Translated histograms given the minimum asymmetry measures.

3.3 Influence of Rigid Transformations and Non-rigid Deformations

One of the important characteristics of the asymmetry descriptors for skin lesion is whether it is robust to the rigid transformations (translation, rotation, scale) and non-rigid deformations. Next we will evaluate their influence to our asymmetry descriptor.

Invariant to Rigid Transformations and Isometric Deformations. GPSs are calculated from graph the Laplace matrix by connecting the neighbour points on object surface in Riemannian space. It has been proved that GPSs are invariant to translation, rotation and isometric deformations [8], because these manipulations do not change the point distance along the surface. Since histogram neglects the spatial information and does not introduce extra isometric deformations, the proposed asymmetry descriptor is translation, rotation and isometric deformation invariant.

For scaling, suppose a ratio γ scaled the manifold M as $M' = \gamma M$. Eigenvalues and eigenvectors of M' can then be calculated as $\lambda'_i = \lambda_i/\gamma^2$ and $\phi'_i = \phi_i/\gamma$. GPSs of the scaled M' proved invariant since $GPS'_i = \phi'_i/\sqrt{\lambda'_i} = (\phi_i/\gamma)/\sqrt{\lambda_i/\gamma^2} = \phi_i/\lambda_i = GPS_i$.

Considering parameter m, l and sign sequence T are scale irrelevant, only descriptor $Des_{i,l}(T)$ in (6) needs checking. Since GPSs are scale invariant, bin location v , representing the quantified value in GPS histograms, stays unchangeable to γ . f representing the frequency counts, is a function of pixel number. As such dividing the pixel number N_l in each segment as (7) can counteracts the influence of γ .

Robustness to Non-isometric Deformations. Since skin lesions are non-rigid objects, they associate with non-isometric deformations most of the time. Besides the

2D shape distortions due to skin surface tension, non- isometric deformations in Z-axis also exist because of the different intensity caused by different dermoscopes.

Given a d-dimensional manifold M and its non-isometric deformation M' , the graph Laplace matrix can be represented as $L = S^{-1}C$ and $L' = S'^{-1}C'$, where S represents the area surrounding by a pixel's neighbour points, and C is a cotangent function [9]. ϕ'_i and λ'_i of the non-isometric perturbation object will remain stable if non-isometric perturbation $\Delta C = |C - C'|$ and $\Delta S = |S - S'|$ are small. Thus the proposed asymmetry descriptor is robust to small non-isometric deformations.

4 Experimental Results

The performance of the asymmetry descriptors for melanoma diagnosis is validated on 311 dermoscopy images [10][11] with resolution ranging from 448×336 to 1098×826 pixels, where there are 88 MMs and 223 BNs. The asymmetry descriptor is a four dimensional feature vector including the minimum asymmetry measures obtained from MI and EI images respectively, as well as their asymmetry measures in the direction perpendicular to the optimal reflectional axis.

$$\text{Asymmetry Features} = \{Asy_{min}(MI), Asy_{min+90^\circ}(MI), Asy_{min}(EI), Asy_{min+90^\circ}(EI)\}$$

4.1 Efficiency of Biology Information

In order to demonstrate the efficiency of the biology information, we compute the asymmetry descriptor in each channel of the conventional RGB images, and compare the diagnosis results with those from the melanin and erythema indexes.

Fig.4 plots the distributions of the probability density functions (pdfs) of the proposed asymmetry descriptors obtained from MI/EI images, as well as those from the conventional RGB images. The asymmetry descriptors in MI/EI images give better separated distributions between MMs and BNs. The application of the biology descriptors increases the discrimination, because MI/EI reflects the pathological condition and the primary cause of the pigmentation inside the lesion. This makes MI and EI ideal descriptors for quantizing of the malignancy of the skin lesions.

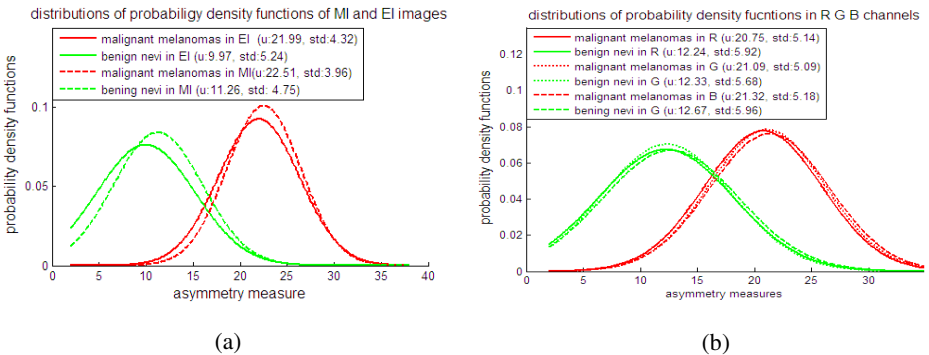


Fig. 4. Distributions of probability density function (pdfs) of the asymmetry descriptors. (a) pdfs in MI and EI images. (b) pdfs in Red, Green and Blue channels in conventional RGB images.

4.2 Robustness to Non-isometric Deformations

To demonstrate the robustness of the proposed asymmetry descriptor to the non-isometric deformations, we calculate the asymmetry descriptors of the same lesion acquired by different dermoscopes. An example of Seborrheic Keratosis is shown in Fig.5. Although the overall shape of the lesion seems unchanged, non-uniform surface tension will always exist, not to mention the different lesion boundaries introduced by the automatic segmentation techniques. Thus 2D shape distortion between two images can be detected from the isolated lesions in Fig.5(ii). Moreover, compared with the original images, colour contrast in two MI images proves similar. Similar contrast with different intensity values only causes a translation along the Z-axis in the pigmentation model, but not greatly distorted the shape. So we assume there are small non-isometric deformations between the post-processed images.

Though both shape and colour distortions are presented, the asymmetry of the first 6 GPSs in Fig.5(ii) appear resembled. This similarity can be also observed from the corresponding histograms. Numerically, asymmetry degree of the lesion in (a) is 8.53, whereas that of the lesion in (b) is 9.27, which gives 8.87% difference.

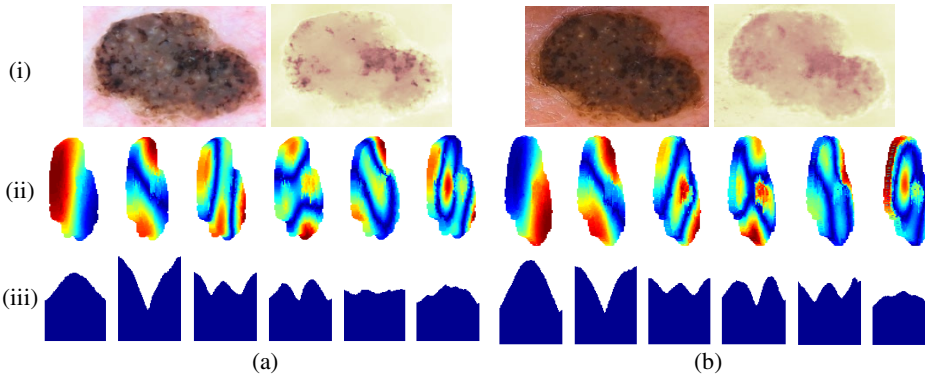


Fig. 5. Asymmetry measure of the same seborrheic keratosis by different dermoscopes. (a) polarized dermoscopy with resolution of 598×492 (b) immersion contact dermoscopy with the resolution of 540×462. (i) Original images and MI images. (ii) The first 6 GPSs with the optimal sign sequence T. (iii) Translated histograms given the minimum asymmetry measures.

4.3 Performance of the Asymmetry Descriptors

In order to demonstrate the efficiency of the proposed asymmetry descriptor, we compute the extrinsic shape and colour asymmetry without GPSs, and compare the classification results with that from the intrinsic descriptors with GPSs. The asymmetry measures without GPSs are defined similarly as the proposed detector. Specifically, a lesion is first segmented into 180 areas. Then each segment is represented by the area proportion of the segment to the whole lesion ($SDes_l$), or by the MI/EI histogram ($CDes_l$). Finally the asymmetry measure without GPSs can be quantified by minimizing the histogram difference in $[0 \pi]$ as (9).

$$SDes_l = \frac{N_l}{N}, \quad CDes_l = \frac{1}{N_l} \sum_{n=1}^{100} f(\text{colour}_l^n) * v(\text{colour}_l^n) \quad (8)$$

$$SAsy = \min \left(\sum_{l=1}^{90} \|SDes_l^t - SDes_l^r\|_2^2 \right), \quad CAsy = \min \left(\sum_{l=1}^{90} \|CDes_l^t - CDes_l^r\|_2^2 \right) \quad (9)$$

Three different classifiers were applied to validate the classification accuracy, including Support Vector Machine (SVM) with radial basis function of a scalar 3, Artificial Neural Networks (ANN) with 1 hidden layer and 5 neurons, and Bayesian classifier (BC) by minimizing the Bayesian risk of the 0/1-loss function. In the training-testing process, 155 dermoscopy images were randomly selected for training and the other half of data were used for testing. For each classification algorithm, we automatically execute the program 30 times and record the average sensitivity, specificity and accuracy as the final results to complement the bias introduced by the inconsistency of the random selection of the training data.

Table.1 shows the classification results of the asymmetry descriptors with and without GPSs from each classifier. The accuracy from the GPSs-based asymmetry descriptor is approximate 5% higher than that of the combination of shape and colour asymmetry descriptors without GPSs. The best diagnosis for the test data are 88.50% sensitivity, 81.92% specificity and 83.26% accuracy employing the SVM classifier.

Table 1. Classification results of the asymmetry descriptors with and without GPSs

	Training (%) (with/without GPSs)			Testing (%) (with/without GPSs)		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
SVM	94.53/93.02	84.07/81.44	87.03/84.72	88.50/83.63	81.92/75.81	83.26/78.02
ANN	88.58/86.36	80.49/77.82	82.78/80.24	86.27/82.92	78.64/74.23	79.96/76.69
BC	90.25/86.40	81.10/76.19	83.69/79.08	83.35/79.56	75.86/70.78	77.98/73.26

5 Conclusions

This paper proposes a novel reflectional asymmetry descriptor by minimizing the histogram difference of the global point signatures on the pigmentation elevation model of the cutaneous lesions. Melanin and Erythema indexes have been proved more efficient than colour intensities in characterizing the pigmentation of the lesions. The proposed asymmetry descriptor is invariant to 2D rigid transformations and robust to non-isometric deformations. Competitive classification results of 88.50% sensitivity and 81.92% specificity have been achieved for melanoma diagnosis.

References

- [1] Jerant, A., Johnson, J., Sheridan, C., Caffrey, T.: Early Detection and Treatment of Skin Cancer. *American Family Physician* 6, 357–368 (2000)
- [2] Stoecker, W., Li, W., Moss, R.: Automatic Detection of Asymmetry in Skin Tumors. *Computerized Medical Imaging and Graphics* 16, 191–197 (1992)
- [3] Seidenari, S., Pellacani, G., Giannetti, A.: Digital Videomicroscopy and Image Analysis with Automatic Classification for Detection of Thin Melanomas. *Melanoma Research* 9, 163–171 (1999)

- [4] d'Amico, M., Ferri, M., Stanganelli, I.: Qualitative Asymmetry Measure for Melanoma Detection. In: IEEE Int. Symposium. Biomedical Imaging: Nano to Macro, pp. 1155–1158. IEEE Press, New York (2004)
- [5] Bae, Y., Nelson, J., Jung, B.: Multimodal Facial Colour Imaging Modality for Objective Analysis of Skin Lesions. *Journal of Biomed Optics* 13(6), 064007 (2008)
- [6] Lee, T., Ng, V., Gallagher, R., Coldman, A., McLean, D.: DullRazor: A software approach to hair removal from images. *Computers in Biology and Medicine* 27, 533–543 (1997)
- [7] Liu, Z., Sun, J., Smith, M., Smith, L., Warr, R.: An automatic mean-shift based segmentation for pigmented skin lesions. *Medical Imaging Understanding and Analysis*, 121–126 (2010)
- [8] Balkan, M., Niyogi, P.: Laplace-beltrami Eigenfunctions for Deformation Invariant Shape Representation. *Neural Computation* 15(6), 1373–1396 (2002)
- [9] Ovsjanikov, M., Sun, J., Guibas, L.: Global Intrinsic Symmetries of Shapes. In: Symposium on Geometry Processing, vol. 27, pp. 1341–1348. Eurographics Association, Aire-la-Ville (2008)
- [10] Home page - Dr. Eric Ehrsam, <http://dermoscopic.blogspot.com/>
- [11] Dermatology Information System, <http://www.dermis.net/>

Identification of Paediatric Tuberculosis from Airway Shape Features

Benjamin Irving¹, Pierre Goussard², Robert Gie², Andrew Todd-Pokropek¹,
and Paul Taylor¹

¹ University College London, London WC1E 6BT, UK
b.irving@ucl.ac.uk

² Tygerberg Hospital, Stellenbosch University,
Western Cape 7530, South Africa

Abstract. Clinical signs of paediatric pulmonary tuberculosis (TB) include stenosis and deformation of the airways. This paper presents two methods to analyse airway shape and detect airway pathology from CT images. Features were extracted using (1) the principal components of the airway surface mesh and (2) branch radius and orientation features. These methods were applied to a dataset of 61 TB and non-TB paediatric patients. Nested cross-validation of the support vector classifier found the sensitivity of detecting TB to be 86% and a specificity of 91% for the first 10 PCA modes while radius based features had a sensitivity of 86% and a specificity of 94%. These methods show the potential of computer assisted detection of TB and other airway pathology from airway shape deformation.

1 Introduction

The prevalence of tuberculosis (TB) remains high in many developing countries while the accuracy of paediatric TB detection is low, and a combination of tests including imaging is used. Automated airway analysis has the potential to improve the detection of airway pathology such as TB. A common sign of primary TB in children is airway deformation caused by lymphadenopathy [1]. This can take the form of displacement and stenosis of airway branches, and widening of the carinal angle [2]. Hila, mediastinal, subcarinal and paratracheal lymph nodes are commonly affected and the most common sites for compression are: the trachea, left main bronchus (LMB), right main bronchus (RMB) and bronchus intermedius (BI) [3]. This sign is more sensitive in children because the airways are more malleable and primary TB tends to affect the lymph nodes. Lymphadenopathy can also indicate other pathology but is useful for detecting TB when used in conjunction with other tests and is likely to indicate TB in areas with a high TB prevalence.

Paediatric airways are considerably smaller than those of adult patients, which means a lower resolution using the same voxel size, and fewer branches can be identified. Movement artefacts are also more likely because it is not possible to perform a breath hold scan on infants [10].

Point distribution models (PDM) are a common method of modelling shape variation. Anatomical landmarks or a mesh are used to represent a shape and the variation in position of corresponding points is calculated. Principal component analysis (PCA) can be applied to the PDM to reduce the dimensionality of the representation, identifying the principal modes of variation. These techniques have been applied successfully in a number of cases including facial morphology [5]. However, very little research has focussed on airway shape modelling. A previous study developing a shape model of the airways focussed on patient specific models and required manual interaction [4].

An alternative and more intuitive approach is to use features that correspond directly to clinical observations but this requires more background knowledge. In this paper we present a complete system for analysing pathological airway shape variation and compare two approaches for identifying TB cases: one using features generated using the principal modes of variation of a surface mesh, and another using features based on the branch radius and orientation. We test both on a dataset of TB and non-TB cases. Contributions of this work include a method to generate a shape model of normal and pathological airway variation (the authors are not aware of any previous method to model airway pathology and particularly to distinguish between TB and non-TB datasets) and methods focussed on paediatric datasets. Additional contributions include a novel method for automatically generating airway landmarks based on the airway topology and centreline, extension of mesh-warping to suit stenosed airway shape variation and the training of a classifier on airway shape data.

2 Method

An automated airway segmentation approach was used and the centreline and bifurcation points extracted. Corresponding landmark points were generated and a template mesh was warped to each airway. A shape model was then developed using the principal modes of the corresponding vertices. Cross-section diameter measurements were made for each branch and used to generate the second feature set.

2.1 Dataset and Airway Extraction

The dataset used in this study consists of TB and non-TB cases. 29 chest CT scans of paediatric patients diagnosed with definite or probable TB from a positive culture, or bronchoscopy and CT findings were acquired from Tygerberg Hospital in South Africa (mean age 22 ± 26 months) and 32 chest CT scans of paediatric patients with a non-TB diagnosis were acquired from Gt Ormond St Hospital, London (mean age 38 ± 22 months). Voxel size in the axial plane ranged from 0.3 - 0.5 mm and slice thickness 0.7 - 1 mm. 13 cases with completely obstructed branches were previously manually excluded from the dataset because these cases can be easily identified and are not of interest for building a shape model. The age difference between the groups is within one standard deviation and age does not influence airway proportions in children [8].

The airways were segmented using an existing method [6]. This method uses morphological closing and reconstruction to enhance possible airway locations in the axial, coronal and sagittal directions. A region growing method, seeded at the trachea, is then used to extract the airway region. The structure of the airways is found using centreline extraction, branch point detection and branch labelling. Palágyi et al. [9]’s skeletonisation method is used for the extraction of the centreline because of its previous application to the airways. This is an iterative thinning approach, where each surface voxel is analysed in terms of orientation and connectivity and simple points are iteratively removed.

False branching can occur because of surface deformation (particularly when pathology is present) and, therefore, branch pruning is required. We found that false branches connected to the trachea, LMB and RMB can be longer than true branches further down the tree, and false branches may bifurcate. Therefore, a multilevel pruning system was developed that removed branches less than a specified length (l) and removed larger false branches associated with the primary branches and smaller false branches associated with later generations. Three pruned trees ($T_{l_1}, T_{l_2}, T_{l_3}$) were created with pruning $l_1 > l_2 > l_3$. A final tree was constructed from T_{l_1} for the trachea, T_{l_2} for the LMB and RMB and from T_{l_3} for the remaining branches. A one voxel thick centre line was used to identify the branching structure, shown in Figure 1: a branch point was defined as a point with three neighbours in the 3x3x3 surrounding region.

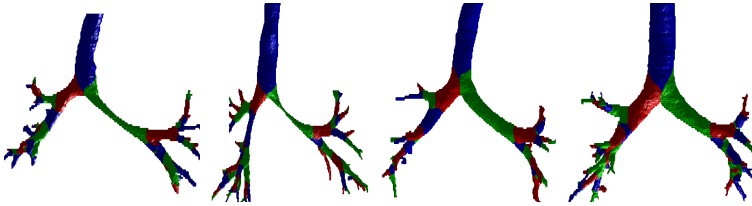


Fig. 1. Paediatric airway segmentation and branch-point identification. The two cases on the left show signs of TB while the others are non-TB cases.

2.2 Corresponding Surface Point Generation and Mesh Alignment

Surface point correspondence is required to derive features from a shape model while diameter based features require only regular sampling of the branch. Branch points are the only major anatomical landmarks and, therefore, corresponding points were generated by calculating the intersection between the surface and vectors orthogonal to the smoothed centreline at equidistant positions along each branch (Figure 2). The generated points take into account branch topology, medial line curvature and surface deformation. The analysis was performed on the trachea, RMB and LMB (commonly deformed by lymphadenopathy).

As discussed earlier, two sets of features are being considered, the principal modes of the surface deformation and branch radius/direction based features.

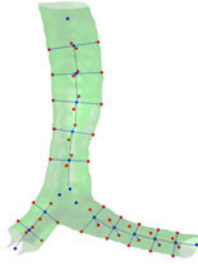


Fig. 2. Surface point placement using the centreline and bifurcation points

The surface points were used to calculate the two orthogonal diameters at each cross section along each branch. These points were generated from 60 equidistant points on the medial line of the trachea, 50 along the LMB and 30 along the RMB. A subset of the corresponding points (generated from 5 equidistant points on the Trachea and LMB, and 2 on the RMB) were used to warp a mesh onto each airway using Thin Plate Spline (TPS) warp. TPS warping is a common method of aligning objects using a set of landmark points [5]. TPS attempts to perform realistic deformation by minimising the bending energy [2]. The TPS function that minimises the energy is:

$$f_j(P_j) = \sum_{i=1}^k w_{ij} U(P_j - P_{ij}) + a_0 + a_x x + a_y y + a_z z \quad (1)$$

where f is the new position of the point and f_j is a component of f , $j \in (x, y, z)$, P are the landmark points on the shape and w_{ij} are the weighting factors. w_{ij} can be found from the corresponding landmark points.

Further matching is required so that the template mesh is aligned with each target mesh (as shown in Figure 3). The simplest method is to project the template mesh to the closest point on the target mesh [5] but this can lead to unrealistic deformation while not covering small deformations. Figure 4 shows this mesh misalignment because of narrow sections caused by stenosis and the proposed solution.

Kaus et al [7] optimise the fit based on the distance between the meshes while an additional force preserves the mesh structure. We add a third term based on surface orientation. For each vertex on the template mesh (t_i), a force ($F_{i,tot}$) is calculated to direct the warp. The closest point (r_i) on the object mesh component is included to align the meshes (Eqn 2) and an internal forcing component is included to preserve the size of the faces (change in the distance of each of the p neighbouring vertices to a vertex t_i from the initial distance v_{0j}) (Eqn 3). An expansion/contraction force is also added, based on the normal of each vertex \hat{n}_i (calculated from the normal of the surrounding faces) controlled by the distance and direction of the target mesh to $F_{i,1}$ (Eqn 4). This improves performance for small surface indentations/protrusions associated with stenosis.

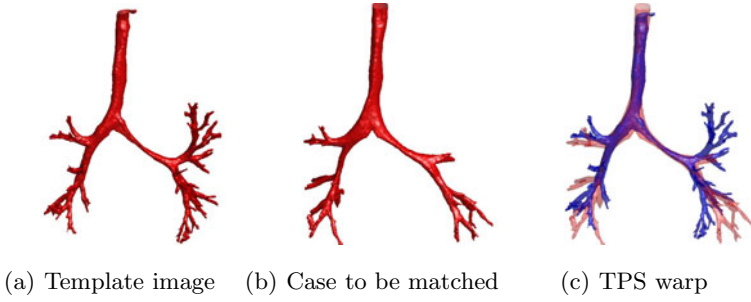


Fig. 3. TPS warp using landmarks on the trachea, LMB and RMB

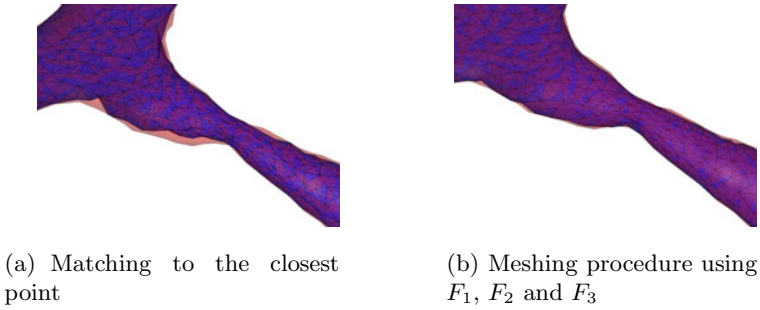


Fig. 4. Mesh matching

$$F_{i,1} = \mathbf{r}_i - \mathbf{t}_i \quad \text{nearest point ext. force} \quad (2)$$

$$F_{i,2} = \sum_j^p \hat{v}_j (\|\mathbf{v}_j\| - \|\mathbf{v}_{0j}\|) \quad \text{where } \mathbf{v}_j = \mathbf{t}_j - \mathbf{t}_i \quad \text{internal force} \quad (3)$$

$$F_{i,3} = \hat{n}_i (\hat{n}_i \cdot F_{i,1}) \quad \text{normal ext. force} \quad (4)$$

$$F_{i,tot} = \alpha F_{i,1} + \beta F_{i,2} + \gamma F_{i,3} \quad (5)$$

In Equation 5, the forces are weighted with α , β and γ . This procedure is applied iteratively until stability is reached.

2.3 Feature Extraction and Classification

Each shape is represented as a $3n$ dimensional vector where n is the number of vertices in the mesh; $n \approx 1500$ was used in this study. Each shape was aligned using Generalised Procrustes analysis and PCA was applied to reduce the dimensionality and obtain a set of features for classification. PCA applies a linear transform that projects the PDM onto an uncorrelated space and can be used to extract relevant features [3]. PCA modes are ordered by the variance and, therefore, can be used to reduce the dimensionality of the feature vector. For PCA, it can be shown that the eigenvectors of the covariance matrix

$\Sigma = XX^T$ (where each column of X is a $3n$ vector for each airway) can be used to project the dataset into the uncorrelated space (b) represented by the eigenvectors $\mathbf{b} = \Phi^T(\mathbf{x} - \bar{\mathbf{x}})$ where the projection matrix (Φ^T) is the transpose of the eigenvector matrix (Φ). Therefore, a measurement vector $\mathbf{x} = \bar{\mathbf{x}} + \Phi\mathbf{b}$ can be represented in terms of the mean and displacement along each mode [3].

Three radius based features were calculated for each branch: the maximum ratio of the orthogonal diameters for each branch ($\max(\frac{d1}{d2})$), the ratio between the branch length and average branch diameter ($\frac{d_{ave}}{l}$) and the maximum ratio of local minima and neighbouring local maxima of the diameter as a function of position on the branch ($\frac{l_{max1} + l_{max2}}{2l_{min}}$). These features, based on advice from our clinical partners, were used as indicators of branch circularity, thickness and local stenosis, and were calculated for the trachea, RMB and LMB. The carinal angle was also calculated for each airway by fitting a line to the first third of the RMB and LMB and calculating the angle from bifurcation. All features were normalised.

Once a set of features was found to represent each airway in the dataset, a classifier was trained to distinguish between TB and non-TB cases. A Support Vector Machine (SVM) was chosen as the classifier because of its suitability for small datasets and the PRtools implementation of SVM was used. Leave-one-out cross validation (LOOCV) and nested CV were used to evaluate the classifier.

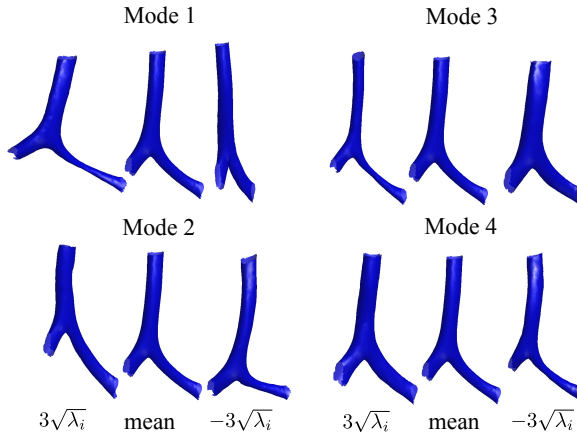
3 Results

Parameters for the mesh warp α , β and γ were determined by comparing the volume generated from both the template mesh and the original mesh ($V_{dif} = (V_{temp} \setminus V_{case}) \cup (V_{case} \setminus V_{temp})$). In order to focus on local errors instead of differences due to mesh face sizes, a morphological closing was applied ($V_{open} = V_{dif} \circ K$ where K is 6-connected kernel) in order to remove 1-voxel thick errors but retain larger local errors. Optimum parameters are around $\alpha = 0.2$ and $\gamma = 1$, where proportion of *general error* (without closing) is less than 0.022 and *local error* (with closing) is less than 0.002. Without the expansion force ($\gamma = 0$) then the minimum errors are 0.05 and 0.02 respectively. Fixed parameters were used for the whole dataset but could be chosen for each individual airway.

The SVM classifier was trained and tested on the two sets of features. Classification using PCA features were performed using the first 10 modes which represented 90% of the shape variation. Figure 5 shows the mean and variation from $-3\sqrt{\lambda_i}$ to $3\sqrt{\lambda_i}$ along the first 4 modes. Classification was also performed on the 10 radius and orientation based features. This classifier was optimised by adjusting the “trade-off parameter” C (between 5 and 500) and the degree of the polynomial kernel (between 1 and 13) while running LOOCV for each choice. These values were chosen to cover a reasonable range of parameters but further optimisation could be performed. LOOCV was used because the dataset was too small to divide into a testing and training set. However, adjusting the SVM parameters with LOOCV allows the best classifier to be selected but can lead to a biased measure of accuracy. Therefore, to determine an unbiased sensitivity

Table 1. Sensitivity and specificity using (1) the PCA and (2) the radius and orientation based feature set with LOOCV and Nested CV

	LOOCV		Nested CV	
	PCA	Rad	PCA	Rad
Sensitivity	93%	93%	86%	86%
Specificity	94%	94%	91%	94%

**Fig. 5.** Variation along the first four PCA modes

and specificity without an independent training set, nested CV was used [11]. Nested CV includes a second LOOCV loop with parameter optimisation inside the full LOOCV loop and results have been shown to be close to that of an independent testing set [11]. Using LOOCV, the classifiers performed the same and parameters of $C=100$ and 3 and polynomial degree of 3 and 1 were found for the PCA and radius based classifiers, respectively (Table 1). The radius based features performed slightly better when tested using nested CV (6 compared to 7 misclassified out of 61). The software was written in Matlab and C++ and tested on a 2.0 GHz quad-core processor. Generation of features from a segmented dataset and cross validation: ≈ 700 s for the PCA based feature vector and ≈ 1200 s for the radius based feature vector.

4 Discussion

In this paper we discuss two methods to quantify and detect airway shape deformation due to TB. Both these methods were able accurately to distinguish between paediatric cases with TB and without TB, and demonstrate the potential of these techniques to assist in the detection of airway pathology. PCA based features may be more generalizable and, more effective for differentiating other types of pathology without adjusting the feature choice.

The datasets were collected from two hospitals and it is possible that population differences also have an effect on the classification. However, the features extracted using PCA correspond to clinical signs of TB. Examining Figure 5, the modes correspond to stenosis and widening of the carinal angle, which is consistent with clinical signs of TB [1]. The other feature set was based on characteristics of airway pathology.

This paper shows the potential of automated airway analysis to assist in the identification of pathology with possible CAD applications. The model could be developed further by training on localised pathology, or applied to other areas such as airway deformation and narrowing caused by congenital cardiac disease.

References

1. Andronikou, S., Wiesenthaler, N.: Modern imaging of tuberculosis in children: thoracic, central nervous system and abdominal tuberculosis. *Pediatr. Radiol.* 34, 861–875 (2004)
2. Bookstein, F.: Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Trans. Pattern Anal. Mach. Intell.* 11, 567–585 (1989)
3. Cootes, T., Taylor, C., Cooper, D., Graham, J., et al.: Active shape models-their training and application. *Comput. Vis. Image. Und.* 61, 38–59 (1995)
4. Deligianni, F., Chung, A., Yang, G.: Nonrigid 2-D/3-D registration for patient specific bronchoscopy simulation with statistical shape modeling: Phantom validation. *IEEE Trans. Med. Imag.* 25, 1462–1471 (2006)
5. Hutton, T., Buxton, B., Hammond, P., Potts, H.: Estimating average growth trajectories in shape-space using kernel smoothing. *IEEE Trans. Med. Imag.* 22, 747–753 (2003)
6. Irving, B., Taylor, P., Todd-Pokropek, A.: 3D segmentation of the airway tree using a morphology based method. In: *Second International Workshop on Pulmonary Image Analysis, MICCAI* (2009)
7. Kaus, M., Pekar, V., Lorenz, C., Truyen, R., Lobregt, S., Weese, J.: Automated 3-D PDM construction from segmented images using deformable models. *IEEE Trans. Med. Imag.* 22, 1005–1013 (2003)
8. Masters, I., Ware, R., Zimmerman, P., Lovell, B., Wootton, R., Francis, P., Chang, A.: Airway sizes and proportions in children quantified by a video-bronchoscopic technique. *BMC Pulmonary Medicine* 6, 5–13 (2006)
9. Palágyi, K., Tschirren, J., Hoffman, E., Sonka, M.: Quantitative analysis of pulmonary airway tree structures. *Comput. Biol. Med.* 36, 974–996 (2006)
10. du Plessis, J., Goussard, P., Andronikou, S., Gie, R., George, R.: Comparing three-dimensional volume-rendered CT images with fiberoptic tracheobronchoscopy in the evaluation of airway compression caused by tuberculous lymphadenopathy in children. *Pediatr. Radiol.* 39, 694–702 (2009)
11. Varma, S., Simon, R.: Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics* 7, 1–8 (2006)

Automatic Region-of-Interest Segmentation and Pathology Detection in Magnetically Guided Capsule Endoscopy

Philip W. Mewes^{1,2}, Dominik Neumann¹, Oleg Licegevic¹, Johannes Simon¹, Aleksandar Lj. Juloski¹, and Elli Angelopoulou²

¹ Siemens AG, Healthcare Sector, Erlangen, Germany

² Pattern Recognition Lab, University of Erlangen-Nürnberg, Erlangen, Germany
philip.mewes@informatik.uni-erlangen.de

Abstract. Magnetically-guided capsule endoscopy (MGCE) was introduced in 2010 as a procedure where a capsule in the stomach is navigated via an external magnetic field. The quality of the examination depends on the operator’s ability to detect aspects of interest in real time. We present a novel two step computer-assisted diagnostic-procedure (CADP) algorithm for indicating gastritis and gastrointestinal bleedings in the stomach during the examination. First, we identify and exclude subregions of bubbles which can interfere with further processing. Then we address the challenge of lesion localization in an environment with changing contrast and lighting conditions. After a contrast-normalized filtering, feature extraction is performed. The proposed algorithm was tested on 300 images of different patients with uniformly distributed occurrences of the target pathologies. We correctly segmented 84.72% of bubble areas. A mean detection rate of 86% for the target pathologies was achieved during a 5-fold leave-one-out cross-validation.

1 Introduction

Background and Purpose of This Work. Endoscopy of the upper gastrointestinal (GI) tract with flexible endoscopes is a standard clinical procedure. The main disadvantages of this procedure are high invasiveness and patient discomfort. Wireless capsule endoscopy (WCE) was introduced in 2001 and is mainly used in the duodenum. The stomach, in comparison, has large surface and volume and can not be reliably examined with an uncontrolled capsule. Endoscopic capsules that can be steered from the outside by means of magnets have been reported in [7][12][11]. In this paper we use human data from the clinical study of [11] that stems from 29 volunteers and 24 patients (Fig. 1c and 1d). A single dataset from one patient contains on average 3600 images. For the MGCE procedure, the patient’s stomach is filled with water and the capsule is navigated from the outside using an external magnetic field. During the examination the operator can control the motion of the capsule so as to obtain a sufficient number of stomach-surface images with diagnostic value. The quality of the examination

depends on the skill of the operator and his ability to detect aspects of interest in real time. We present a novel two-step computer-assisted diagnostic-procedure (CADP) algorithm for detecting, during the examination, two distinct stomach pathologies: gastritis and haematin. Gastritis is one of the main causes of stomach cancer. Chronic gastritis usually appears as a reddish and blotched turgor, while acute gastritis often appears as a small circle-shaped bleeding (Fig. 2b). Haematin on the other hand, accrues from blood coming in contact with stomach acid. It is a sign of gastrointestinal bleeding and is an indicator of many significant diseases. It appears as a uniform brownish shape close to the stomach wall (Fig. 2a). The proposed algorithm is divided into two steps: 1) a region-of-interest (ROI) segmentation to separate medically relevant sections of the image from parts containing bubbles; 2) a contrast-normalized filtering to identify and localize possible lesions of pathologies. For this 2nd task we develop a feature vector, which is used for classifying pathologies in a machine learning approach.

State of the Art. In a typical WCE examination a large number of frames is medically irrelevant, as they either do not show pathologies or contain mainly intestinal juices, bubbles or debris. To assist the physician in reviewing up to ten hours of video material, software for computer-aided diagnosis (CAD) has been developed. In [2,14,13] different descriptors for the task of blood and ulcera detection and topographic segmentation of the GI tract are investigated. Topographic segmentation is addressed in [1] and [5] for the purpose of a more efficient and faster review. In [4] and [15] the issue of eliminating redundant frames, as well as those with intestinal juices, is addressed. In [6] a set of color and texture based features for the detection of intestinal bleedings is presented. These methods are not directly applicable to MGCE. The duodenum, when compared to the stomach, exhibits different pathologies and imaging conditions, such as texture and distance to objects of interest. In [8] a method for the automatic detection of gastritis aspects in MGCE was presented. In this paper we used the method from [8] as a starting point and developed a ROI pre-segmentation for bubbles and an improved segmentation method for different pathologies (gastritis and haematin). Compared to the aforementioned existing CAD algorithms which are only used for review, our method indicates the pathologies during the examination itself. The operator can, for instance, navigate to a suspicious region for further closer inspection. Even if 100% accuracy is not achieved, our algorithm may still point out lesions which would otherwise be missed.

2 Automatic Pathology Detection in MGCE

2.1 Region of Interest Segmentation

Endoscopic images obtained via MGCE may contain bubbles and mucus (Fig. 1a). Such a region within an image usually contains no medically relevant information. One should, thus, segment such regions to exclude them from further processing. In [15] and [10] two methods for automatic detection of intestinal juices are presented that exclude entire frames containing intestinal juices and bubbles for the review process. The location of bubbles in the stomach is relatively stable

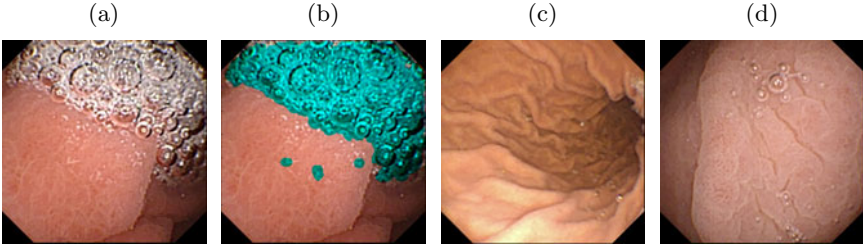


Fig. 1. (a) sample image with bubbles. (b) segmented bubble area (c) lesser curvature and body (d) mucosa close-up B.

and localized. Rejecting an entire frame with bubbles could therefore eliminate images of whole stomach regions together with possible pathologies.

The segmentation of bubble regions is not as straightforward as it may initially appear. The edges of most bubbles appear bright, but the translucent part is dominated by the color of the underlying tissue (Fig. 1(a)). Therefore, an intensity based approach is not effective. One has to combine geometry and color cues in order to detect the entire bubbles region. To that end, we investigated a large variety of robust feature descriptors initialized with different key point detectors. We randomly chose as our training data 100 healthy and 100 diseased images from our dataset containing different amounts and spatial distributions of bubbles. Regions with bubbles were hand-labeled. We denote one pixel within an image as a tuple (x, y, v) , where x and y are the pixel coordinates, and v is a pixel value. An image \mathcal{I} (432×432 pixels) is defined as a set of pixels, $\mathcal{I} = \{(x_1, y_1, v_1), \dots, (x_N, y_N, v_N)\}$ where N is the total number of pixels in the image. We define $\mathcal{O}_1 \subseteq \mathcal{I}$ as the set of all pixels in a bubble area and $\mathcal{O}_2 \subseteq \mathcal{I}$ as the area without bubbles. In addition $\mathcal{O}_1 \cap \mathcal{O}_2 = \emptyset$, and $\mathcal{O}_1 \cup \mathcal{O}_2 = \mathcal{I}$.

We considered 5 descriptors (SIFT, steerable filters, GLOH, SPIN, Cross Correlation), each of which can be initialized with one of 5 different key point detectors (Hessian-Laplacian, Harris-Laplacian, Harris-Affine, Hessian-affine, Harris)¹. For each combination of descriptor type and key point detector we obtained n descriptors (feature vectors) $D_k, k = 1, \dots, n$, which were calculated on the corresponding image patches $P_k \subset \mathcal{I}, k = 1, \dots, n$ around the detected key points. We used the labeled data to train (using Adaboost) a 2-class classifier that could distinguish between descriptors corresponding to \mathcal{O}_1 and \mathcal{O}_2 . We evaluated each descriptor / key-point-detector combination for bubble detection using a 5-fold leave-one-out cross-validation (LOO CV).

Training phase: An image patch P_k is characterized by its center point $c = (x, y)$ and an ellipse centered at c which is defined by one or two radii $r_{1,2}$ respectively. For the supervised training input, a label

¹ Affine covariant features [Online]. Available: <http://www.robots.ox.ac.uk/~vgg/research/affine/>

$$l(P_k) = \begin{cases} 0 & \text{if } P_k \subset \mathcal{O}_1 \\ 1 & \text{if } P_k \subset \mathcal{O}_2 \end{cases} \quad (1)$$

was assigned. The remaining P_k s (i.e. image patches which do not completely belong to either of the classes) were discarded from training. For training the feature vector D_k was extended by 10 features based on HSV and RGB histograms. More specifically, the mean, variance, skew, kurtosis and entropy were computed for the S and R channel histograms for each patch P_k . These features enhance the classifier’s ability to distinguish structures with geometric similarities to bubbles (such as a round shape) which are, however, part of the stomach mucosa or a target pathology.

Testing: After the training was completed, the new labels $l(P_k)$ were computed based on the obtained classifier. We define two measures for the quality of segmentation:

$$E(\mathcal{O}_i) = \frac{\text{card}\left(\left(\bigcup_{\{P_k|l(P_k)=1\}} P_k\right) \cap \mathcal{O}_i\right)}{\text{card}(\mathcal{O}_i)} \quad (2)$$

with $i = \{1, 2\}$ ² $E(\mathcal{O}_1)$ measures the proportion of the ground-truth bubble area covered by image patches classified as $l = 1$, while $1 - E(\mathcal{O}_2)$ measures the area without bubbles, that is wrongly covered by image patches classified as $l = 1$. We found that the combination of a Hessian-affine key point detector and a steerable filter descriptor yields the best results (see Table I (first row)). We refer to the union $\{\cup(P_k|l(P_k) = 0)\}$ as the binary mask \mathcal{I}_b and to bubble free area as $\mathcal{I}_{bn} = \mathcal{I} \setminus \mathcal{I}_b$. Note that, the use of an elliptic shape for the image patches P_k yields small areas between the image patches which are not classified as bubbles area. To overcome this problem we introduced a circular morphologic dilation element with a radius r_s . The circle radius adapts to the size of the surrounding image patches to ensure gap closing between large image patches and prevent the dilation of non-bubbles areas. A neighborhood around the dilation element is defined as a window \mathcal{W} of 30×30 pixels around the center of the structuring element. The radius r_s at a position x, y within \mathcal{I}_b is computed from the average radii of all $P_k|P_k \subseteq \mathcal{W}$.

The improvement obtained with the dilation can be seen in Table II. The quantitative measurements were obtained by equation 2. After the segmentation we checked if some pathologies were wrongly classified as bubbles and confirmed that this was not the case (Visual results in Fig. IIb).

2.2 Contrast Normalization and Region Localization

For the following steps only \mathcal{I}_{bn} (i.e. areas without bubbles) are considered. The pathological lesions exhibit low intensity while the surrounding tissue has high intensity values. In order to detect such lesions a LoG-edge detection, as proposed by [8], is performed. Thereby, \mathcal{I}_{bn} is convolved with a Laplacian-of-Gaussian

² $\text{card}(\cdot)$ denotes the cardinality, i.e. the number of pixels in the image subset.

Table 1. Results of Bubble Segmentation

Method	$E(O_1)$	$1 - E(O_2)$
Direct use of descriptor	81.53%	4.24%
With adaptive morphological operation	84.72%	5.74%

kernel \mathcal{K} . The resulting image \mathcal{I}_c is subsequently searched for prominent edges whose magnitude is above a certain threshold. Because images suffer from varying contrast and lighting conditions, a contrast normalized variable threshold is introduced. To that end, \mathcal{I}_c is converted into a probability mass function $f(b_i)$ with N_{b_i} intervals. The contrast normalized threshold t_c is then computed as follows. We first define the index t_{pos} of an interval b_i

$$t_{pos} = \sum_{i=N_{b_i}}^1 I \left(\left(\sum_{j=N_{b_i}}^i f(b_j) \right) < \left(\frac{\sum f(b_i)}{4} \right) \right) \quad (3)$$

where I is an indicator function which is equal to 1 if its argument is true and zero otherwise. t_c is then computed as $t_c = \max(b_{t_{pos}})$.

A binary edge image $\mathcal{I}(t_c)$ is then computed using the variable threshold t_c . $\mathcal{I}(t_c)$ is 1 for $\mathcal{I}_c \geq t_c$ and 0 otherwise. Given such a binary image our goal is to merge connected pixels into areas representing possible locations of pathologies. A morphologic closing operator with a disc-shaped structuring element of a radius of 5 pixels is applied to $\mathcal{I}(t_c)$. Subsequently, using 4-connectivity, pixels are grouped into structures that we refer to as S_j with $j = 1, \dots, m$ where m is the total number of structures per image. On average, images contain 45 S_j 's with a mean size of 20×14 pixels. Ultimately, we want to classify each structure S_j into one of the three classes: the *gastritis class* $\mathcal{C}1$, the *haematin class* $\mathcal{C}2$ and the *negative class* $\mathcal{C}3$ without any abnormalities. For training purposes, all S_j are superimposed on to the original RGB-image. Visual inspection shows that all possible lesions of $\mathcal{C}1$ and $\mathcal{C}2$ are detected by the above described region localization method. All computed S_j are therefore directly hand-labeled by an expert so that a label L is assigned to each structure S_j

$$L(S_j) = \begin{cases} 0 & \text{if } S_j \in \mathcal{C}1 \text{ (gastritis class)} \\ 1 & \text{if } S_j \in \mathcal{C}2 \text{ (haematin class)} \\ 2 & \text{if } S_j \in \mathcal{C}3 \text{ (negative class)} \end{cases} \quad (4)$$

2.3 Feature Extraction

The following sets of features are extracted for each structure S_j (Fig. 2c).

Geometric features (GF). Geometric features describe the specific shape of pathologies. We extract the maximal vertical and horizontal dimensions of S_j and

Table 2. Number of features computed for each S_j

Feature Group	Feature Extractor	No. of Features	Composition of Feature Vector
GF	Aspect ratio of S_j	1	width/height
	dimensions of S_j	2	width,height
	bounding box fill factor	1	%
	Hu moments	7	invariant moments
TF	lbp H	5	local binary pattern on three channels of HSV colorspace
	lbp S	5	
	lbp V	5	
HF	RGB	15	Mean, variance, skew kurtosis and entropy of each channel Histogram
	HSV	15	
	CIE	15	
Total no. of features		71	

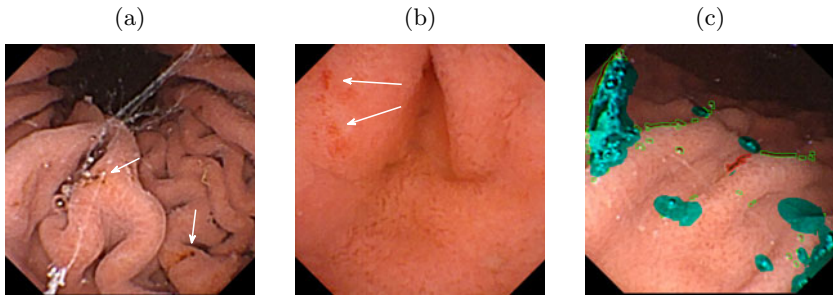


Fig. 2. Sample images of target pathologies. (a): haematin. (b): gastritis lesions. (c): Computed structure S_j (green border). S_j with red border contains a gastritis lesion. Segmented bubble area in turquoise.

of its minimal bounding box and their corresponding aspect ratio. Furthermore the ratio between the area covered by S_j and the area of the minimal bounding box is computed. We refer to this feature as *fill factor*. For further geometric analysis we also extract the Hu moments [3] for each S_j .

Texture features (TF). To further investigate textures within each S_j local binary patterns (lbp) [9] are computed for each channel of the HSV colorspace. The resulting lbp histogram is computed for each S_j .

Histogram features (HF). Color based features aim to distinguish between color characteristics of structures containing gastritis and structures of class $\mathcal{C}3$. Thus, we compute the mean, variance, skew, kurtosis and entropy of each histogram-channel of the RGB, HSV and CIE colorspace. The entire collection of features is summarized in Table 2.

Table 3. Results of Pathology Detection using a SVM classifier (in %). Results of AdaBoost in parentheses.

Predicted \ Actual	Healthy Image (C1)	Gastritis Image (C2)	Haematin Image (C3)	Gastritis and Haematin C2/C2
Healthy Image C1	92 (95)	6(5)	2(0)	0(0)
Gastritis C2	8(11)	72(55)	6(12)	14(22)
Haematin C3	7(11)	7(12)	76(70)	10(7)

3 Experiments and Results

To evaluate the automatic detection of pathologies we used 300 images stemming from 44 patients. The images were analyzed by a medical expert and each detected structure S_j was manually labeled. The dataset of 300 images consisted of 100 images for each of the two target pathologies and a set of 100 healthy images. For the supervised learning approach we compared results of the Adaboost and SVM classifiers implemented in a 5-fold LOO CV. Most images contained more than a single structure S_j with or without a pathology. An image was attributed to one of the pathology classes if at least one structure S_j was classified as pathologic. An image was counted as healthy if none of the structures S_j was classified as pathologic. The best classification results were obtained using a SVM classifier³. A confusion matrix was obtained for each image test-set of the 5-fold LOO CV. The average confusion matrix of all 5 runs can be seen in Table 3. Out of 100 images with gastritis, 86% were correctly classified (Table 3). From the 100 images containing haematin, 86% were correctly classified. However, we also detected an average of 12% of pathological images, which exhibited both, haematin and gastritis. Finally out of the 100 healthy images 92% were correctly classified.

4 Discussion and Conclusion

Our experiments have shown that our computer-assisted diagnostic-procedure algorithm can be used for indicating gastritis and gastrointestinal bleedings in MGCE. The presented algorithm includes a preprocessing step that discards areas with bubbles. This step is crucial for all following image processing steps in the presented method and may have implications for the development of future applications on this imaging modality. Pre-segmentation performs accurately in detecting areas with bubbles without hiding pathologies or large amounts of non-bubble tissue areas. Based on the results from the above pre-processing step, a method was presented that is able to automatically detect two kinds of pathologies on MGCE images. We achieved sensitivity and specificity results well over 80% for healthy and diseased images. The moderate sensitivity between the two target pathologies is due to the similarity between gastritis and haematin

³ LIBSVM [Online]. Available: <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>

aspects in terms of color and texture. Still, the algorithm performs well, especially within the context of real-time warnings for certain gastritis aspects.

References

1. Berens, J., Mackiewicz, M., Bell, D.: Stomach, intestine, and colon tissue discriminators for wireless capsule endoscopy images. In: Fitzpatrick, J., Reinhardt, J. (eds.) *Proceedings of SPIE*, vol. 5747, p. 283 (2005)
2. Cunha, J., Coimbra, M., Campos, P., Soares, J.: Automated topographic segmentation and transit time estimation in endoscopic capsule exams. *IEEE T. Med. Imaging* 27(1), 19–27 (2007)
3. Hu, M.: Visual pattern recognition by moment invariants. *IRE Trans. Info. Theory* 8, 179–187 (1962)
4. Karargyris, A., et al.: A video-frame based registration using segmentation and graph connectivity for Wireless Capsule Endoscopy. In: *Life Science Systems and Applications Workshop LiSSA 2009, IEEE/NIH*, pp. 74–79. IEEE, Los Alamitos (2009)
5. Mackiewicz, M., Berens, J., Fisher, M.: Wireless capsule endoscopy color video segmentation. *IEEE T. Med. Imaging* 27(12), 1769–1781 (2008)
6. Mackiewicz, M., Fisher, M., Jamieson, C.: Bleeding detection in wireless capsule endoscopy using adaptive colour histogram model and support vector classification. In: Joseph, M., Josien, P. (eds.) *Proceedings of SPIE*, vol. 6914, p. 69140R (2008)
7. Menciasci, A., Valdastrì, P., Quaglia, C., Buselli, E., Dario, P.: Wireless steering mechanism with magnetic actuation for an endoscopic capsule. In: *Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2009*, pp. 1204–1207. IEEE, Los Alamitos (2009)
8. Mewes, P., Neumann, D., Juloski, A., Angelopoulou, E., Hornegger, J.: On-the-fly detection of images with gastritis aspects in magnetically-guided capsule endoscopy. In: Ronald, M., van Ginneken, B. (eds.) *Medical Imaging - Computer-Aided Diagnosis. Proceedings of the SPIE*, vol. 7963, p. 79631I3. SPIE, San Jose (2011)
9. Ojala, T., Pietikäinen, M., Harwood, D.: A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition* 29(1), 51–59 (1996)
10. Reingruber, H.: Intestinal content detection in capsule endoscopy using robust features @ONLINE (2009), <http://www.recercat.net/handle/2072/43221>
11. Rey, J., Ogata, H., Hosoe, N., Ohtsuka, K., Ogata, N., Ikeda, K., Aihara, H., Pangtay, I., Hibi, T., Kudo, S., Tajiri, H.: Feasibility of stomach exploration with a guided capsule endoscope. *Endoscopy* 42(7), 541–545 (2010)
12. Swain, P., Toor, A., Volke, F., Keller, J., Gerber, J., Rabinovitz, E., Rothstein, R.: Remote magnetic manipulation of a wireless capsule endoscope in the esophagus and stomach of humans. *Gastrointest Endosc* (2010)
13. Szczypinski, P., Klepaczko, A.: Selecting texture discriminative descriptors of capsule endoscopy images. In: Zinterhof, P. (ed.) *Proceedings of ISPA 2009*, pp. 701–706. IEEE, Los Alamitos (2009)
14. Szczypinski, P., Sriram, P., Sriram, R., Reddy, D.: Model of deformable rings for aiding the wireless capsule endoscopy video interpretation and reporting. *Comp. Imag. Vis.*, 167–172 (2006)
15. Vilarino, F., Spyridonos, P., Pujol, O., Vitria, J., Radeva, P., De Iorio, F.: Automatic detection of intestinal juices in wireless capsule video endoscopy. In: Tang, Y., Wang, S., Lorette, G., Yeung, D., Yan, H. (eds.) *ICPR 2006*, vol. 4, pp. 719–722. IEEE, Los Alamitos (2006)

Simultaneous Segmentation and Grading of Hippocampus for Patient Classification with Alzheimer's Disease

Pierrick Coupé¹, Simon F. Eskildsen¹, José V. Manjón², Vladimir Fonov¹, D. Louis Collins¹, and The Alzheimer's Disease Neuroimaging Initiative*

¹ McConnell Brain Imaging Centre, Montreal Neurological Institute, McGill University, Montreal, Canada. University, 3801 University Street, Montreal, Canada H3A 2B4

² Instituto de Aplicaciones de las Tecnologías de la Información y de las Comunicaciones Avanzadas (ITACA), Universidad Politécnica de Valencia, Camino de Vera s/n, 46022 Valencia, Spain

Abstract. Purpose: To propose an innovative approach to better detect Alzheimer's Disease (AD) based on a finer detection of hippocampus (HC) atrophy patterns. Method: In this paper, we propose a new approach to simultaneously perform segmentation and grading of the HC to better capture the patterns of pathology occurring during AD. Based on a patch-based framework, the novel proposed grading measure estimates the similarity of the patch surrounding the voxel under study with all the patches present in different training populations. The training library used during our experiments was composed by 2 populations, 50 Cognitively Normal subjects (CN) and 50 patients with AD. Tests were completed in a leave-one-out framework. Results: First, the evaluation of HC segmentation accuracy yielded a Dice's Kappa of 0.88 for CN and 0.84 for AD. Second, the proposed HC grading enables detection of AD with a success rate of 89%. Finally, a comparison of several biomarkers was investigated using a linear discriminant analysis. Conclusion: Using the volume and the grade of the HC at the same time resulted in an efficient patient classification with a success rate of 90%.

Keywords: hippocampus segmentation, hippocampus grading, patient classification, nonlocal means estimator, Alzheimer's disease.

1 Introduction

The atrophy of medial temporal lobe structures, such as the hippocampus (HC) and entorhinal cortex, is potentially specific and may serve as early biomarkers of

* Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (www.loni.ucla.edu/ADNI). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. ADNI investigators include (complete listing available at www.loni.ucla.edu/ADNI/Collaboration/ADNI_Authorship_list.pdf).

Alzheimer's disease (AD) [1]. In particular, the atrophy of the HC can be used as a marker of AD progression since changes in HC are closely related to changes in cognitive performance of the subject [1]. The evaluation of HC atrophy is usually estimated by volumetric studies on anatomical MRI, requiring a segmentation step that can be very time consuming when done manually. This limitation can be overcome by using automatic segmentation methods [2-5]. However, despite the high segmentation accuracy of these HC segmentation approaches, using only the HC volume enables a separation between AD and cognitively normal (CN) subjects with a success rate around 72-74% [6]. This limited capability to classify AD patients by using the HC volume only may be due to a simplification of the complex hippocampal atrophy patterns to a volume changing measurement. Recently, several shape analysis methods have been proposed [7-8] to capture detailed patterns of change in order to obtain a more accurate classification. These approaches provide a slightly better classification rate of around 77% [6].

Inspired by work in image denoising [9], a new nonlocal patch-based label fusion method has recently been proposed to segment anatomical structures [5]. By taking advantage of the redundancy of information present within the subject's image, as well as the redundancy across the training subjects, the patch-based nonlocal means scheme enables robust use of a large number of samples during estimation. In [5], this approach has been applied to label fusion for the segmentation of anatomical structures. We propose an extension of this patch-based segmentation method in order to evaluate the similarity (in the nonlocal means sense) of the intensity content of one MRI compared to several training populations. By using training populations with different pathological status (e.g., CN subjects and patients with AD), a nonlocal means estimator is used to evaluate the proximity (i.e., the grade or the degree of atrophy in case of AD) of each voxel of the MRI under study compared to the training populations. Since the grade estimation and the label fusion steps require the same patch comparison, simultaneous segmentation and grading of HC can be achieved in one pass. In the proposed approach, the nonlocal patch-based comparison is used to efficiently fuse the HC segmentations of MRI in a training database and at the same to aggregate the pathological status of the populations constituting the training database. Finally, the average grading value obtained over the segmented HC is proposed as a new biomarker to estimate the pathological status of the subject under study. The contributions of the paper are: *i*) the introduction of an innovative approach to better characterize the patterns of pathology (e.g., atrophy) in AD through the new concept of HC grading, *ii*) the presentation of a method to automatically and simultaneously perform the segmentation and the grading of HC, and *iii*) the demonstration that the proposed approach can be used as a novel biomarker to efficiently achieve patient classification in the context of AD.

2 Materials and Methods

2.1 Dataset and Preprocessing

In this study, the ADNI database (www.loni.ucla.edu/ADNI) was used to validate the proposed approach. This database contains both 1.5T and 3.0T T1-w MRI scans. For

our experiments, we randomly selected 120 MRI scans, 60 1.5T MRI baseline scans of CN subjects and 60 1.5T MRI baseline scans of patients with AD. All the selected images were preprocessed as follows: 1) correction of inhomogeneities using N3 [10], 2) registration to the stereotaxic space using a linear transform to the ICBM152 template (1x1x1 mm³ voxel size) [11] and 3) cross-normalization of the MRI intensity using the method proposed in [12]. After preprocessing, all the MRIs are coarsely aligned (linear registration), tissue intensities are homogeneous within each MRI volume (inhomogeneity correction) and across the training database (intensity normalization). From the 120 processed MRI scans, 20 scans (10 CN and 10 AD) were randomly selected to be used as **seed dataset**. The left and right hippocampi of this **seed dataset** were then manually segmented by an expert at our centre. The manual segmentations of the **seed dataset** were propagated to the 100 remaining scans constituting our **test dataset**. After segmentation propagation using [5], the **test dataset** was composed of 100 MRI (50 CN subjects and 50 patients with AD) with their corresponding automatic segmentations.

2.2 Method Overview

In nonlocal means-based approaches [9], the patch $P(x_i)$ surrounding the voxel x_i under study is compared with all the patches $P(x_j)$ of the image Ω whatever their spatial distance to $P(x_i)$ (it is the meaning of the term “nonlocal”). According to the patch similarity between $P(x_i)$ and $P(x_j)$, estimating by the Sum Squared Difference (SSD) measure, each patch receives a weight $w(x_i, x_j)$:

$$w(x_i, x_j) = e^{-\frac{\|P(x_i) - P(x_j)\|_2^2}{h^2}}$$

where $\|\cdot\|_2$ is the L2-norm computed between each intensity of the elements of the patches $P(x_i)$ and $P(x_{s,j})$, and h is the smoothing parameter of the weighting function. This weighting function is designed to give a weight close to 1 when the SSD is close to zero and a weight close to zero with the SSD is high. Finally, all the intensities $u(x_j)$ of the central voxels of the patches $P(x_j)$ are aggregated through a weighted average using the weights $w(x_i, x_j)$. In this way, the denoised intensity $\hat{u}(x_i)$ of the voxel x_i can be efficiently estimated:

$$\hat{u}(x_i) = \frac{\sum_{j \in \Omega} w(x_i, x_j) u(x_j)}{\sum_{j \in \Omega} w(x_i, x_j)}$$

In [5], we introduced this estimator in the context of segmentation by averaging labels instead of intensities. By using a training library of N subjects, whose segmentations of structures are known, the weighted label fusion is estimated as follows:

$$v(x_i) = \frac{\sum_{s=1}^N \sum_{j \in \Omega} w(x_i, x_{s,j}) l(x_{s,j})}{\sum_{s=1}^N \sum_{j \in \Omega} w(x_i, x_{s,j})}$$

where $l(x_{s,j})$ is the label (i.e., 0 for background and 1 for structure) given by the expert to the voxel $x_{s,j}$ at location j in training subject s . It has been shown that the nonlocal means estimator $v(x_i)$ provides a robust estimation of the expected label at x_i [5]. With

a label set of $\{0,1\}$, voxels with value $v(x_i) \geq 0.5$ are considered as belonging to HC and the remaining voxels as background.

In this paper, we propose to extend it to efficiently aggregate pathological status in order to estimate the proximity (in the nonlocal means sense) of each voxel compared to both populations constituting the training library. To do that, we introduce the new concept of patch-based grading that reflects the similarity of the patch surrounding the voxel under study with all the patches present in the different training populations. In this way, the neighborhood information is used to robustly drive the search of anatomical patterns that are specific to a given subset of the training library. When the training populations include data from subsets of patients with different stages of the pathology progression, this approach provides an estimation of the grade (i.e., degree of atrophy in case of AD) for each voxel:

$$g(x_i) = \frac{\sum_{s=1}^N \sum_{j \in \Omega} w(x_i, x_{s,j}) \cdot p_s}{\sum_{s=1}^N \sum_{j \in \Omega} w(x_i, x_{s,j})}$$

where p_s is the pathological status of the training subject s . In our case, $p_s = -1$ was used for AD status and $p_s = 1$ for CN status. A negative grading value (respectively, a positive grading value) $g(x_i)$ indicates that the neighborhood surrounding x_i is more characteristic of AD than CN (respectively, of CN than AD). The absolute value $|g(x_i)|$ provides the confidence given to the grade estimation. When $|g(x_i)|$ is close to zero, the method indicates that the patch under study is similarly present in both populations and thus is not specific to one of the compared populations and provides little discriminatory information. When $|g(x_i)|$ is close to 1, the method detects a high proximity of the patch under study with the patches present in one of the training population and not in the other. Finally, for each subject, an average grading value is computed over all voxels in the estimated segmentation (i.e., for all x_i with $v(x_i) \geq 0.5$) by assigning the same weight to the left and right HC (i.e., $\bar{g} = (\bar{g}_{left} + \bar{g}_{right}) / 2$). During all our experiments, the default parameters proposed in [5] have been used. The patch size was fixed to $7 \times 7 \times 7$ voxels and the search window of similar patches has been limited within a restricted volume of $9 \times 9 \times 9$ voxels for computational reasons (i.e., Ω is replaced by a cubic volume V_i centered on x_i). Finally, the smoothing parameter h^2 was locally set as the minimal SSD found between the patch under study and all the patches in the training library as proposed in [5].

2.3 Validation Framework

Segmentation accuracy validation: In order to evaluate the segmentation accuracy of the method proposed in [5] on patients with AD, we first perform a leave-one-out procedure on the 20 subjects with manual segmentation composing the **seed dataset**. The $N=16$ closest training subjects (in the SSD sense, see [5] for details) were equally selected within both populations (i.e., 8 AD and 8 CN). The Dice's kappa was then computed by comparing the expert-based segmentation, used as gold standard, and the segmentation obtained automatically. This first validation is used to support the fact that the segmentation propagation over the 100 subjects in our **test dataset** from the 20 subjects in our **seed dataset** is done in an accurate manner.

Grading validation: After the segmentation propagation step, a leave-one-out procedure is performed over the 100 subjects of the **test dataset**. For each subject, the N closest training subjects are selected equally in both populations. This is done to ensure that the size of the “*patch pool*” from AD population is coarsely similar to the size of the “*patch pool*” from CN population. To save computational time, N is automatically adjusted according to the obtained \bar{g} . In the first iteration, $N=20$ (10 CN and 10 AD). If the resulting $|\bar{g}| < 0.1$ (i.e., the confidence in the obtained grade is low), the size of the used training library is increased by 20 to $N=40$ (20 CN and 20 AD). This process is repeated until $|\bar{g}| > 0.1$ or $N > 80$. The sign of the final grading value is used to estimate the pathological status of the testing subjects. Finally, the success rate of the patient classification is provided to demonstrate the robustness of the proposed new biomarker.

Comparison of biomarkers for patient classification with AD: The last part of our validation framework is the comparison of two biomarkers (HC volume and HC grade) and the investigation of their combination. The segmentations obtained at the same time as the grading were used to obtain the HC volume for each of the subjects in the **test dataset**. Through a leave-one-out procedure, each subject was classified by using optimal boundary separating both populations. This optimal boundary was obtained by performing a linear discriminant analysis over the 99 remaining subjects. This approach was applied to volume-based classification, grade-based classification and the combination of both volume and grade. The success rate (SR), the specificity (SPE), the sensitivity (SEN), the positive predictive value (PPV) and negative predictive value (NPV) are presented for each of the tested biomarkers (see [6] for details on these quality metrics).

3 Results

Table 1 shows the segmentation accuracy obtained on the **seed dataset** by using $N=16$ training subjects (8 CN and 8 AD). For the CN population, the median Dice’s Kappa was similar to the Dice’s Kappa presented in [5] on healthy young subjects from the ICBM database, which demonstrates the robustness of the segmentation method. A lower median Dice’s Kappa value was obtained for the AD population. A median Dice’s Kappa value superior to 0.8 indicates a high correlation between manual and automatic segmentations, and a median Dice’s Kappa value superior 0.88 is similar to the highest published values in literature [3-4]. The difference between populations might come from two sources. First, the higher anatomy variability of patients with AD makes the segmentation more difficult and may require a larger training library. Second, the smaller HC volumes of patients with AD, due to the HC atrophy, can negatively bias the Dice’s Kappa index measure. Finally, these results indicate that a similar accuracy can be expected during the segmentation propagation step to the 100 subjects of the **test dataset**.

Table 1. Median Dice's Kappa values (with the standard deviation) obtained on the **seed dataset** composed of 20 MRI (10 CN and 10 AD) with manual segmentations.

<i>Median Dice's Kappa (standard deviation)</i>	<i>Left HC</i>	<i>Right HC</i>	<i>Both HC</i>
CN population	0.891 (0.035)	0.866 (0.038)	0.883 (0.037)
AD population	0.830 (0.042)	0.858 (0.035)	0.838 (0.038)

Figure 1 shows the final grading values for the 100 subjects of the **test dataset**. In the perfect case, the 50 first subjects (CN) should have positive average grading values and the 50 last (AD) should have negative average grading values. As shown in the graph, the success rate of the classification was 89% (5 false positive CN and 6 false negative AD). Figure 1 also presents the size of the used training library for each of the testing subjects. Most of the test subjects were classified by using only $N=20$ training subjects. Around 5% of test subjects seem to require larger training library (i.e., $N>80$) since at the end of the procedure $|\bar{g}|$ is still inferior to 0.1.

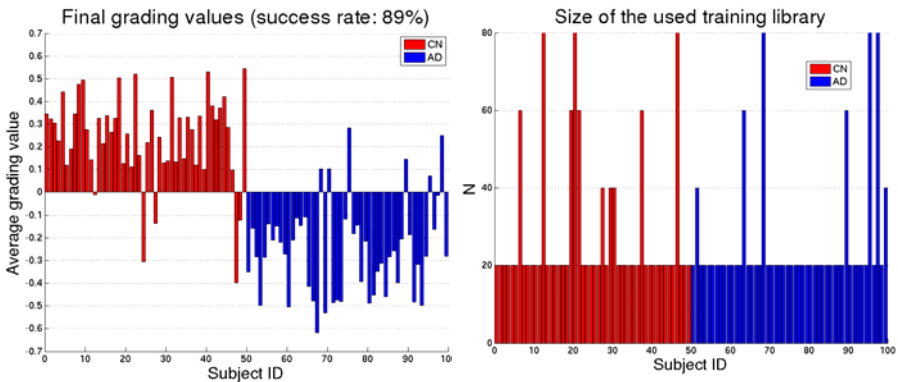
**Fig. 1.** Left: the final average grading values obtained for the **test dataset**. Right: the used size of training library (i.e., N) for all the testing subjects.

Figure 2 shows the grading maps obtained for 2 test subjects (1 CN and 1 AD). The corresponding average grading values and the estimated volumes are also provided for left and right HC. While the volume of HC is similar for these 2 subjects, and thus does not allow an efficient patient classification, their grading values provide a useful indication on their pathological status. Visually, the CN subject clearly appears closer to the CN population (mainly red color related to values close to 1) while the AD patient is visually closer to the AD population (mainly purple and black colors related to values close to -1). Finally, Fig. 2 also provides a visual assessment of the quality of the segmentation propagation on the **test dataset**. For a given subject, the segmentation and the grading maps were obtained in less than 5 minutes using a single core of an Intel Core 2 Quad Q6600 processor at 2.4 GHz with $N=20$.

Table 2 presents the results of the patient classification for the different biomarkers under consideration. These results clearly demonstrate the advantage of using the grading approach (89% of success rate) compared to the classical volumetric approach (78% of success rate). The SEN, SPE, PPV and NPV obtained by our grading approach were higher than the ten methods compared in [6] involving Voxel-Based Morphometry (VBM), cortical thickness, HC volume and HC shape. The higher SR of our volumetric approach compared to the results presented in [6] might come from differences in the test dataset used here or due to a higher accuracy and consistency of the segmentation method used compared to [2]. It is also interesting to note that the optimal boundaries found by linear discriminant analysis provided similar results as using 0 as threshold value as in the previous experiment (see Fig 1.). Finally, using the volume and the average grade of the HC simultaneously provides a very high success rate of 90%.

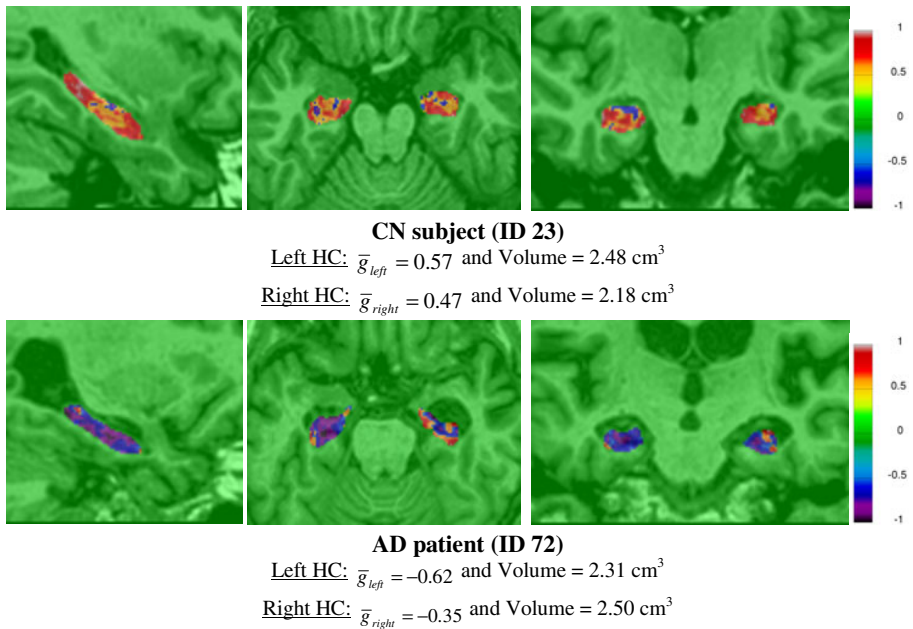


Fig. 2. Top: the obtained grading map for one CN subject (ID 23). Bottom: the obtained grading map for one AD patient (ID 72). The slices of both subjects have the same position in the stereotaxic space. Red color indicates a grading close to 1 (i.e., CN) and black color indicates a grading close to -1 (i.e., AD).

Table 2. Results of the patient classification (AD vs CN) for the different biomarkers under investigation. These results were obtained by using linear discriminant analysis through a leave-one-out procedure on the test dataset.

<i>AD vs. CN</i>	<i>SR</i>	<i>SEN</i>	<i>SPE</i>	<i>PPV</i>	<i>NPV</i>
HC volume	78%	72%	84%	82%	75%
HC grading	89%	86%	92%	91%	87%
HC volume and grading	90%	88%	92%	92%	88%

4 Conclusion

In this paper, a new method is proposed to robustly detect the hippocampal atrophy patterns accruing during AD. Based on a nonlocal means estimation framework, the proposed novel grading measure (i.e., the atrophy degree in AD context) enables an accurate distinction between CN subjects and patients with AD leading to a success rate of 89% when used alone, and 90% when combined with HC volume. These results are competitive compared to the AD detection performance of VBM, cortical thickness, HC volume and HC shape methods extensively compared in [6]. In contrast to these approaches, our method has the advantage of simplicity (it can be coded in few hundred lines of code), low computational cost (does not required non-rigid registration), robustness of the process (all the subjects get final grading maps) and the possibility to achieve *individual* classifications based on a single time point contrary to *group* classification or longitudinal studies. These first results are promising and indicate that the new HC grading approach could be a useful biomarker to efficiently detect AD. Further work will investigate the possibility to discriminate population of patients with Mild Cognitive Impairment (MCI) compared to AD or CN.

References

1. Frisoni, G.B., Fox, N.C., Jack, C.R., Scheltens, P., Thompson, P.M.: The clinical use of structural MRI in Alzheimer disease. *Nature Reviews Neurology* 6(2), 67–77 (2010)
2. Chupin, M., Hammers, A., Liu, R.S., Colliot, O., Burdett, J., Bardinet, E., Duncan, J.S., Garnero, L., Lemieux, L.: Automatic segmentation of the hippocampus and the amygdala driven by hybrid constraints: method and validation. *Neuroimage* 46(3), 749–761 (2009)
3. Collins, D.L., Pruessner, J.C.: Towards accurate, automatic segmentation of the hippocampus and amygdala from MRI by augmenting ANIMAL with a template library and label fusion. *Neuroimage* 52(4), 1355–1366 (2010)
4. Lotjonen, J.M., Wolz, R., Koikkalainen, J.R., Thurfjell, L., Waldemar, G., Soininen, H., Rueckert, D.: Fast and robust multi-atlas segmentation of brain magnetic resonance images. *Neuroimage* 49(3), 2352–2365 (2010)
5. Coupe, P., Manjon, J.V., Fonov, V., Pruessner, J., Robles, M., Collins, D.L.: Patch-based segmentation using expert priors: application to hippocampus and ventricle segmentation. *Neuroimage* 54(2), 940–954 (2011)
6. Cuingnet, R., Gerardin, E., Tessieras, J., Auzias, G., Lehericy, S., Habert, M.O., Chupin, M., Benali, H., Colliot, O.: Automatic classification of patients with Alzheimer’s disease from structural MRI: A comparison of ten methods using the ADNI database. *Neuroimage* (2010)
7. Gerardin, E., Chetelat, G., Chupin, M., Cuingnet, R., Desgranges, B., Kim, H.S., Niethammer, M., Dubois, B., Lehericy, S., Garnero, L., Eustache, F., Colliot, O.: Multidimensional classification of hippocampal shape features discriminates Alzheimer’s disease and mild cognitive impairment from normal aging. *Neuroimage* 47(4), 1476–1486 (2009)
8. Csernansky, J.G., Wang, L., Swank, J., Miller, J.P., Gado, M., McKeel, D., Miller, M.I., Morris, J.C.: Preclinical detection of Alzheimer’s disease: hippocampal shape and volume predict dementia onset in the elderly. *Neuroimage* 25(3), 783–792 (2005)

9. Coupe, P., Yger, P., Prima, S., Hellier, P., Kervrann, C., Barillot, C.: An optimized blockwise nonlocal means denoising filter for 3-D magnetic resonance images. *IEEE Trans. Med. Imaging* 27(4), 425–441 (2008)
10. Sled, J.G., Zijdenbos, A.P., Evans, A.C.: A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Trans. Med. Imaging* 17(1), 87–97 (1998)
11. Collins, D.L., Holmes, C.J., Peters, T.M., Evans, A.C.: Automatic 3-D model-based neuroanatomical segmentation. *Human Brain Mapping* 3(3), 190–208 (1995)
12. Nyul, L.G., Udupa, J.K.: Standardizing the MR image intensity scales: making MR intensities have tissue specific meaning. *Medical Imaging 2000: Image Display and Visualization* 1(21), 496–504 (2000)

Localization of the Lumbar Discs Using Machine Learning and Exact Probabilistic Inference

Ayşe Betül Oktay and Yusuf Sinan Akgül

GIT Vision Lab

Department of Computer Engineering, Gebze Institute of Technology
Kocaeli, 41400, Turkey

{oktay, akgul}@bilmuh.gyte.edu.tr

<http://vision.gyte.edu.tr>

Abstract. We propose a novel fully automatic approach to localize the lumbar intervertebral discs in MR images with PHOG based SVM and a probabilistic graphical model. At the local level, our method assigns a score to each pixel in target image that indicates whether it is a disc center or not. At the global level, we define a chain-like graphical model that represents the lumbar intervertebral discs and we use an exact inference algorithm to localize the discs. Our main contributions are the employment of the SVM with the PHOG based descriptor which is robust against variations of the discs and a graphical model that reflects the linear nature of the vertebral column. Our inference algorithm runs in polynomial time and produces globally optimal results. The developed system is validated on a real spine MRI dataset and the final localization results are favorable compared to the results reported in the literature.

Keywords: lumbar disc detection, graphical models, exact probabilistic inference, object detection.

1 Introduction

There are many intervertebral disc and vertebra localization methods for the vertebral column in the literature [9, 14, 5]. Schmidt et al. [13] introduce a probabilistic inference method that measures the possible locations of the intervertebral discs. The approach uses a part-based model that describes the disc appearances by employing a tree classifier [8]. Similar to [13], Alomari et al. [1] use a graphical model that assumes local and global levels with latent variables. The inference on their graphical model is based on the expectation maximization method which is an approximate and iterative inference technique.

This paper introduces a novel method for the automatic localization and labeling of the lumbar discs from T1-weighted sagittal MR images. At the local level, the method uses recent machine learning methods to locally search the disc positions by employing Pyramidal Histogram of Oriented Gradients (PHOG) [4] with Support Vector Machines (SVM). At the global level, the method takes advantage of the chain-like structure of the spine by assuming latent variables for

the disc positions. A polynomial time exact inference method on the chain-like graphical structure [3] is employed to find the final lumbar disc positions and labeling.

Our method has two important advantages. First, our graphical model does not directly use any image intensity information which varies greatly. Instead, we use the PHOG based SVM detector results as the observed data in our model. Also, many local and semi-global features can be conveniently incorporated into the SVM without complicating the global level inference mechanism.

For the second major advantage of our system, we exploit the one dimensional nature of the spine and form a chain like graphical model that contains only the latent variables for the disc positions. The literature includes very efficient polynomial time inference algorithms for chain-like graphical models that we employ to find robust disc locations even for pathological cases. In addition, the chain-like structure of our model makes it possible to include extra information for the end discs of the lumbar region, which are the most problematic sections in terms of localization. As a result, our model does not assume any image dependent spatial disc positions and hence it does not require manual initialization.

The rest of this paper is organized as follows. The disc scoring with PHOG based SVM is described in Section 2. Section 3 includes the chain-like probabilistic graphical model and exact inference. The validation of the method is presented in Section 4. Finally, we conclude in Section 5.

2 Disc Scoring with PHOG Based SVM

The lumbar discs vary in the size, location, shape, and appearance because of pathologies and individual variations. We observe that the most invariant property of a disc image is its edges and orientation rather than the intensity, location, and shape which are used in the disc detection methods in the literature [1,5]. Therefore, we use the PHOG descriptors for feature extraction which are more robust to disc and vertebrae abnormalities.

In the PHOG descriptor extraction, a shape is represented by a histogram of gradient orientations which are quantized into a number of bins. Each bin shows the number of edges that have orientations within a specific angular range. PHOG combines the local image shape [6] with the spatial pyramid kernel [7].

The gradient values are first computed and their orientations are calculated. Then, a PHOG descriptor is calculated for each candidate region. Extracting the PHOG descriptors in the original images (512x512) has high computational cost. Therefore, we use the integral histogram technique [12] to speed up the feature extraction process.

After the extraction of the PHOG descriptors, they are trained with Sequential Minimal Optimization [10] for SVM. We use a sliding window approach for the scoring process. The pixel in the center of the window is assigned a score that indicates if the window contains a disc. These scores are generated by fitting a logistic regression model to the outputs of SVM [11].

Let $d = \{d_1, d_2, \dots, d_6\}$ denote the labels for the lumbar intervertebral discs (T12-L1, L1-L2, L2-L3, L3-L4, L4-L5, and L5-S1 which are shown in Figure 1).

Each disc d_i is trained and scored separately. Therefore for a given image I , each pixel is assigned 6 scores, one for each disc. For further processing, the pixels that have scores less than 0.5 are eliminated.

Consider $p_{d_i}(l_k)$ as the score assigned to the pixel at location l_k for the disc d_i by SVM. For the disc d_i , the assigned scores are denoted as

$$p^i = \{p_{d_i}(l_1), p_{d_i}(l_2) \dots, p_{d_i}(l_m)\}, \quad (1)$$

where $1 < i < 6$ and m is the number of pixels in the image. The next section explains how we use these scores in the inference algorithm.

3 Disc Center Localization with the Graphical Model

We propose a probabilistic graphical model for the final localization of the disc centers. The local features of the discs are captured by the disc scoring process with PHOG and SVM, however it is not sufficient to discriminate the discs in the lumbar region. More contextual global information like positional and orientational differences between the discs and distance to the spinal cord are necessary for the final localization of discs on the spine. We use a chain-like graphical model that combines such global information with the local information gathered from the disc scoring process.

Our graphical model is a chain consisting of 6 nodes and 5 edges where each node represents a lumbar disc (Figure 2). Let $x_k \in \mathbb{R}^2$ be a random variable that assigns node k to its image location. The optimal configuration $x' = \{x_1, x_2 \dots, x_6\}$ assigns all discs $d = \{d_1, d_2 \dots, d_6\}$ to their exact locations.

Our objective is finding the optimal localizations of the disc centers with the maximum a posteriori (MAP) estimate

$$x' = \arg \max_x P(x|p, \alpha), \quad (2)$$

where α represents the parameters learned from the training set and $p = \{p^1, \dots, p^6\}$ are the assigned scores with Eq. 1. $P(x|p, \alpha)$ captures the probability of being a disc and the relation with the neighboring discs. The Gibbs distribution of $P(x|p, \alpha)$ is defined as

$$P(x|p, \alpha) = \frac{1}{Z} \exp \left\{ - \left[\sum \psi_L(I, x_k) + \lambda \sum \psi_G(x_k, x_{k+1}, \alpha) \right] \right\}, \quad (3)$$

where the potential function $\psi_L(I, x_k)$ carries local information about the discs and the potential function $\psi_G(x_k, x_{k+1}, \alpha)$ includes more global information like the distance and orientation. λ is a weighting parameter which is selected as 0.5.

In the local potential function $\psi_L(I, x_k)$, we directly use the scores $p^i = \{p_{d_i}(l_1), p_{d_i}(l_2) \dots, p_{d_i}(l_m)\}$ generated with Eq. 1.

The global potential function $\psi_G(x_k, x_{k+1}, \alpha)$ is defined as

$$\psi_G(x_k, x_{k+1}, \alpha) = U(x_k, x_{k+1}, \alpha) R(x_k, x_{k+1}, \alpha) D(x_k, \alpha), \quad (4)$$

where the functions U and R capture the positional and orientational differences between neighboring disc variables x_k and x_{k+1} , respectively, and $D(x_k, \alpha)$ captures the Euclidean distance to the spinal cord whose position is calculated using the method of [2].

Let $y^i = \{y_1, y_2, \dots, y_t\}$ be the Euclidean distances between the disc variables x_i and x_{i+1} of the training set used in the scoring process where t is the number of samples in the training set. The distance function $U(x_k, x_{k+1}, \alpha)$ is

$$U(x_k, x_{k+1}, \alpha) = \begin{cases} \frac{|E - \mu(y^k)|}{w}, & \text{if } E \in [\min(y^k) - \sigma, \max(y^k) + \sigma] \\ \infty, & \text{else,} \end{cases} \quad (5)$$

where $\mu(y^k)$ is the mean of y^k measured from the training set, $w = \max(y^k) - \min(y^k)$, σ is a threshold, and E is the Euclidean distance between the disc variables x_k and x_{k+1} .

The angular differences between the discs define the curve-like shape of the lumbar region. In order to handle this information, we use the angle information between the discs. Let $r^i = \{r_1, r_2, \dots, r_t\}$ be the angles between the neighboring disc variables x_i and x_{i+1} measured from the training set. The orientation term $R(x_k, x_{k+1}, \alpha)$ is defined as

$$R(x_k, x_{k+1}, \alpha) = \begin{cases} \frac{|O - \mu(r^k)|}{f}, & \text{if } O \in [\min(r^k) - \tau, \max(r^k) + \tau] \\ \infty, & \text{else,} \end{cases} \quad (6)$$

where O gives the angle differences between the disc variables x_k and x_{k+1} , $\mu(r^k)$ is the mean orientation learned from the training set, τ is a threshold, and $f = \max(r^k) - \min(r^k)$.



Fig. 1. An uncropped T1-weighted MRI mid-sagittal view of the lumbar vertebrae and intervertebral discs

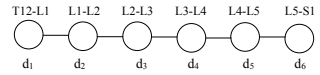


Fig. 2. Our graphical model

3.1 Exact Inference on the Chain

Given a target image I , our objective is to infer the optimal configuration x' by maximizing the Equation [2]. In our graphical model, a node is conditionally

dependent to its neighboring nodes. Instead of using all of the pixels in the image I , we only use the candidate disc centers for the disc localization process. The search space for the optimal x' is decreased by conditional dependence and detected candidates. This allows us to find a globally optimal solution with recursive message passing based dynamic programming [3].

The computation time of our inference method is much better than the inference method of [13] which is based on A* search algorithm. Although A* search can produce the globally optimal solution with an appropriate heuristic, it is potentially intractable. The alternative inference method of expectation maximization [1] is computationally cheaper, but it may not find the globally optimal solution and has known initialization problem.

4 Experimental Results

The developed system is validated on a real MRI dataset for the lumbar spinal column. A 3D MRI volume is 512x512x12 voxels in size. The sagittal view includes the 6 discs of the lumbar vertebrae and some discs from the thoracic vertebrae and sacrum. In order to improve our localization results for the end-lumbar discs, we extended our graphical model with the T11-T12 and S1-S2 discs. However, we do not report their localization results because our purpose is localizing the lumbar discs only.

The dataset consists of MRI volumes of 40 different subjects where 4 of the subjects are pathology-free and the remaining 36 subjects have pathologies like disc degeneration, herniation, and scoliosis, etc. There are totally 240 lumbar intervertebral discs in the dataset and 97 of them have pathologies.

The disc detection is performed on T1-weighted sagittal slices. We use the mid-sagittal slice for both disc detection and disc center localization as it is commonly used in the medical practice. We asked an expert to mark the image region and center point of each disc for the verification and training.

4.1 Disc Scoring Results

For the disc scoring (Section 2), we perform a subset of leave-10-out cross validation. We randomly divide the dataset into 4 subsets each containing 10 subjects. In each sub-experiment, 30 MRI slices from 3 subsets are trained and 10 slices in the other subset are tested. We perform totally 4 sub-experiments, so each slice in the dataset is tested once.

For the SVM training, 90 positive samples and 900 negative samples are used for each disc. For scoring, the features are extracted from the target images. The window size is selected between the minimum and maximum window size in the training set for that disc.

The average SVM classification rates for the testing of 4 subsets are shown in the Table 1. Normally, we produce scores for the windows with logistic model fitting to SVM. If a window is assigned a probability value greater than 0.5, it is evaluated as a detected disc and otherwise it is evaluated as not a disc. The

disc classification is considered as correct if the expert marked disc is completely contained within the window. The classification rate is the number of correctly classified windows over the number of all windows. The misclassified samples are mainly caused by the confusion between the neighboring discs. The borders of the neighboring discs have nearly same orientations, so this causes the detection of the neighboring disc as the target disc most of the times. Note that, the confusion between the neighboring discs are expected because disc scoring is a completely local process.

The detection results reported by [13] has an average classification rate of 70.16%. Our average classification rate is 97% and it shows the robustness of using modified PHOG detector in the disc detection. Note that, the method of [13] and our method are tested on different datasets and [13] uses image sizes of 512x1024 pixels which contain the whole spine.

4.2 Disc Localization Results

Our disc center localization method with graphical model runs on the scoring results of the SVM. The same training and test subsets of the disc scoring process are used for the disc localization.

In order to evaluate the disc localization performance of our system, we use two different methods. In the first method, a disc is evaluated as correctly located if the localized disc center is inside the disc contour. The average disc localization accuracy of our system is 95.42%. The end discs (L4-L5 and L5-S1) have higher error rates than other discs because their positions vary greatly and pathologies generally exist in these discs. The system of Alomari et al. [1] uses 512x512 lumbar region images and it reports an accuracy of 90.7% which is lower than our accuracy. Note that, the method of [1] and our method are not tested on the same dataset.

Table 1. The average classification rates (%) of our system with SVM

Disc detection rates(%)							
	T12-L1	L1-L2	L2-L3	L3-L4	L4-L5	L5-S1	
S1	98.04	97.20	97.85	98.45	97.06	97.99	
S2	97.44	97.08	97.73	98.17	98.08	98.88	
S3	96.41	96.50	96.71	97.89	96.25	98.58	
S4	97.12	97.57	97.78	98.89	97.93	98.96	
Avg	97.25	97.09	97.52	98.35	97.33	98.59	

Table 2. The mean Euclidean distances to ground truth for each subset

Mean Euclidean distances (mm) for each subset							
	T12-L1	L1-L2	L2-L3	L3-L4	L4-L5	L5-S1	
S1	2.06	2.08	3.03	3.89	3.14	4.10	
S2	2.77	2.12	2.40	2.75	4.58	3.85	
S3	3.28	2.78	2.60	2.67	4.11	4.07	
S4	2.29	2.18	1.75	2.72	3.24	2.52	
Avg	2.60	2.29	2.45	3.01	3.77	3.63	

The second evaluation method for the disc center localization is the Euclidean distance to the center labeled by an expert. The mean of the Euclidean distances to the ground truth disc centers are shown in Table 2. The plot box of the Euclidean distances for the lumbar vertebrae connected discs are also shown in Figure 3. In the box plot, the centerline of the box is the median, the top and bottom of lines of the box are 25th and 75th percentiles and the pluses are the statistical outliers. Figure 4 shows the median of the Euclidean distances of our

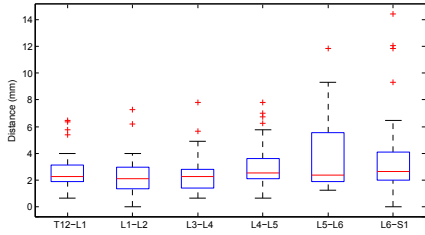


Fig. 3. The box plot shows the Euclidean distances (in mm) to the disc centers

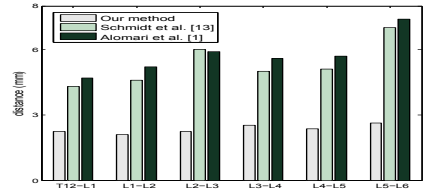


Fig. 4. Median errors of our system, the method of [13] and [1]

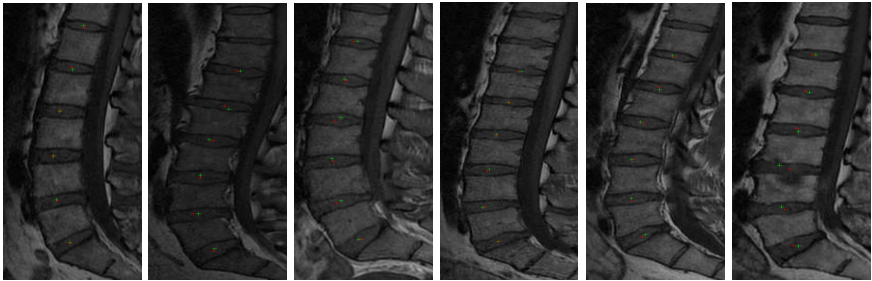


Fig. 5. The results of the disc localization. Red pluses are the ground truth and green pluses are the disc centers localized by our system. The images are cropped for better visualization.

method, the method of [1] and [13]. Our median of the Euclidean distances are always lower than the medians of the other methods [1] and [13]. Note that, [1] and [13] are tested on different datasets. We also show a few visual results of the localizations from our system (Figure 5). It is obvious that our method can localize the lumbar discs in pathological cases.

5 Conclusions

We presented a lumbar intervertebral disc localization method by employing PHOG based SVM and exact inference on a probabilistic graphical model. Our disc scoring method is more robust than other methods in the literature as verified by the experiments. In addition, other local information about the discs such as T2-weighted image features can be easily incorporated into the system. We rely on the robustness of our scoring process to eliminate some disc positions before they enter the dynamic programming, which makes our system more efficient. Also, our dynamic programming based inference mechanism can locate the disc centers in polynomial time without requiring manual initialization. Finally, the method can be extended to localize the whole intervertebral discs of

the whole spine. Our future work includes handling of the missing detected discs and providing scale independency.

Acknowledgments. The authors would like to thank Alpaslan Yildiz for his valuable ideas and Dr. Naci Dumlu from the Giresun State Hospital for providing MRIs.

References

1. Alomari, R.S., Corso, J.J., Chaudhary, V.: Labeling of Lumbar Discs Using Both Pixel- and Object-Level Features With a Two-Level Probabilistic Model. *IEEE Trans. Med. Imaging* 30, 1–10 (2011)
2. Bhole, C., Kompalli, S., Chaudhary, V.: Context-sensitive Labeling of Spinal Structures in MRI Images. In: *SPIE Medical Imaging* (2009)
3. Bishop, C.M.: *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer, Heidelberg (2007)
4. Bosch, A., Zisserman, A., Munoz, X.: Representing Shape with a Spatial Pyramid Kernel. In: *Proceedings of the International Conference on Image and Video Retrieval* (2007)
5. Carballido-Gamio, J., Belongie, S.J., Majumdar, S.: Normalized Cuts in 3-D for Spinal MRI Segmentation. *IEEE Trans. Med. Imaging* 23, 36–44 (2004)
6. Dalal, N., Triggs, B.: Histograms of Oriented Gradients for Human Detection. In: *International Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 886–893 (June 2005)
7. Lazebnik, S., Schmid, C., Ponce, J.: Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories. In: *IEEE Conference on Computer Vision and Pattern Recognition* (2006)
8. Lepetit, V., Fua, P.: Keypoint Recognition Using Randomized Trees. *IEEE Trans. Pattern Anal. Mach. Intell.* 28, 1465–1479 (2006)
9. Peng, Z., Zhong, J., Wee, W., Huei Lee, J.: Automated Vertebra Detection and Segmentation from the Whole Spine MR Images. In: *Conf. Proc. IEEE Engineering in Medicine and Biology Society* (2005)
10. Platt, J.C.: *Advances in Kernel Methods*, pp. 185–208. MIT Press, Cambridge (1999)
11. Platt, J.C.: *Advances in Large Margin Classifiers*, pp. 61–74 (1999)
12. Porikli, F.: Integral Histogram: A Fast Way To Extract Histograms in Cartesian Spaces. In: *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 829–836 (2005)
13. Schmidt, S., Bergholdt, M., Dries, S., Schnorr, C.: Spine Detection and Labeling Using a Parts-Based Graphical Model. In: Karssemeijer, N., Lelieveldt, B. (eds.) *IPMI 2007*. LNCS, vol. 4584, pp. 122–133. Springer, Heidelberg (2007)
14. Seifert, S., Wachter, I., Schmelzle, G., Dillmann, R.: A Knowledge-based Approach to Soft Tissue Reconstruction of the Cervical Spine. *IEEE Trans. Med. Imaging* 28(4), 494–507 (2009)

Automatic Contrast Phase Estimation in CT Volumes

Michal Sofka¹, Dijia Wu¹, Michael Sühling¹, David Liu¹, Christian Tietjen²,
Grzegorz Soza², and S. Kevin Zhou¹

¹ Image Analytics and Informatics, Siemens Corporate Research, Princeton, NJ, USA

² Computed Tomography, Siemens Healthcare, Forchheim, Germany

Abstract. We propose an automatic algorithm for phase labeling that relies on the intensity changes in anatomical regions due to the contrast agent propagation. The regions (specified by aorta, vena cava, liver, and kidneys) are first detected by a robust learning-based discriminative algorithm. The intensities inside each region are then used in multi-class LogitBoost classifiers to independently estimate the contrast phase. Each classifier forms a node in a decision tree which is used to obtain the final phase label. Combining independent classification from multiple regions in a tree has the advantage when one of the region detectors fail or when the phase training example database is imbalanced. We show on a dataset of 1016 volumes that the system correctly classifies native phase in 96.2% of the cases, hepatic dominant phase (92.2%), hepatic venous phase (96.7%), and equilibrium phase (86.4%) in 7 seconds on average.

1 Introduction

Computed Tomography (CT) remains the most common modality used in the imaging of the liver and for the diagnosis of focal liver lesions. Multiphase study obtained by multidetector-row CT (MDCT) during defined circulatory phases best outlines the vasculature and improves detection and characterization of parenchyma lesions [10]. Automatic algorithms for lesion classification, segmentation, and serial comparison require the knowledge of the contrast phase to get the most accurate results (Figure 1). Currently, the phase information either needs to be entered manually or extracted from Dicom tags. Unfortunately, these entries are often incorrect or missing (15-20% of the cases in our experience, see Section 4), and their format varies across hospitals and clinicians.

In this paper, we present an algorithm for automatic contrast phase classification based on the image intensity in local regions. Since several organs and anatomical structures are enhanced differently

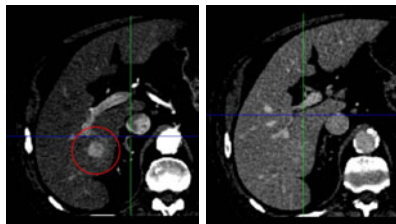


Fig. 1. Contrast phase estimation is important for automatic liver lesion detection and segmentation. The hyperdense liver lesion is clear in the arterial phase (left), but almost invisible in the venous phase (right).

during a specific contrast phase, the algorithm exploits these different levels of enhancement to automatically determine the phase label. Our approach starts by a robust learning-based detector of anatomical structures. The local regions surrounding the structures are then used to train a classifier identifying the following contrast phases: native (NP), hepatic arterial dominant (HADP), hepatic venous (HVP), and equilibrium (EP). The phases are explained in detail in Section 2.

The proposed algorithm uses only the image content to automatically detect the contrast phase without relying on the Dicom tags which often do not contain any label information. This way, the phase labeling is standardized rather than subjective and it can be used in the automatic contrast-specific scan analysis such as lesion detection, segmentation, and comparison in the follow-up studies [2,6]. Since these algorithms might be inaccurate or even fail when the contrast information is incorrect, the requirements on the phase detection accuracy are high. Moreover, since the contrast estimation is adding computation to the overall scan analysis, the algorithm must be efficient.

The robustness of our algorithm comes from the joint analysis of several local image regions surrounding anatomical structures. Discriminative learning-based anatomy detectors (PBT [9]) are trained using a large annotated database of images. The anatomical structures that we use are the liver, aorta, vena cava at the level of renal artery bifurcation, and kidney center (see Figure 4 for examples). The learning algorithm solves the inherent ambiguity of the anatomy locations (e.g. along the vessel or center of the liver) and approaches performance of the human annotations. Multiple image regions used in the phase classification increases robustness where the phase cannot be determined using one region alone and when one of the anatomy detectors fail.

The proposed phase estimation technique is novel in three aspects. First, it presents a way to effectively combine evidence from multiple image regions using confidence of anatomical detectors. Second, the phase classification procedure is based on a decision tree, where each node is a multi-class LogitBoost classifier. Third, the organization of the tree allows to leverage prior knowledge such as the effectiveness of each landmark to classify particular phase or to adjust the tree topology based on the amount of training labels for each phase. We will show on a database of CT scans that the technique can accurately classify the phase in 93% of the cases.

2 Contrast Enhancement in CT

In the first breath-hold of approximately 24 seconds, the arterial dominant phase (HADP) acquires images with arterial and arteriportal enhancement [5]. In the second breath-hold of about 10 seconds, portal enhancement is scanned beginning 60 seconds after the injection of contrast medium (hepatic venous phase, HVP). During this phase, maximum enhancement of the liver parenchyma and strong enhancement of renal cortex and medulla is achieved. Finally, a 3-min delayed scan (equilibrium phase, EP) is acquired. The renal calices and pelvis

start to fill with excreted contrast after approximately 120 seconds [4]. Native phase (NP) scan is obtained without contrast injection. Specific enhancement of organs in each of the phases makes it possible to estimate the contrast phase based on the organ scans. Often, only one phase is acquired (see Section 4).

In the clinical routine, contrast phase information is usually added manually to the series description or image comments Dicom fields. Unfortunately, this information is not structured or standardized. The new Dicom objects aimed to capture timing and phase of enhancement [1] are not widely adopted and there already exists a vast amount of data with unstructured tags. The acquisition timing in the image meta data (if available) could be used to extract time delay between multiple scans. However, the delay after the start of contrast injection could not be obtained since the power injector for the agent is not coupled with the scanner. Moreover, all phases are not always scanned to reduce the amount of radiation to the patient. In our approach, we do not rely on the meta data but rather use the image regions surrounding liver, aorta, vena cava (at the renal vein branching), and kidneys to automatically estimate the contrast phase.

To illustrate the image intensity changes in various contrast phases, we computed the following statistics. Each anatomical region was used to obtain a histogram of intensities and a statistical value of the histogram was plotted for all volumes (Figure 2). We used the mean value of the histogram for aorta and vena cava and peak value for liver and kidneys (due to intensity inhomogeneities). The figure shows, that the regions are indicative of the contrast phase. However, it is not possible to classify all the phases by one of the regions alone as evidenced by overlaps in the plots.

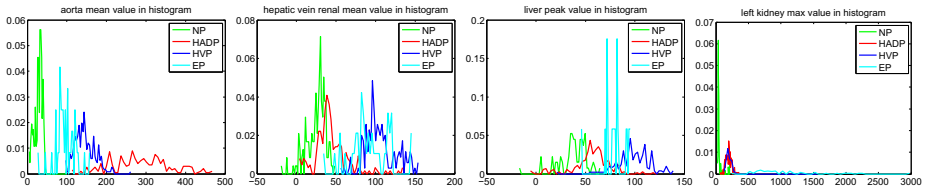


Fig. 2. Intensity statistics computed from 1016 volumes within the image regions defined by aorta, vena cava, liver, and left kidney (right kidney plot is similar). The intensities within the selected anatomical structures are indicative of the contrast phase.

3 Algorithm

Our algorithm automatically determines a probabilistic estimate of the image contrast phase. The set of phases \mathbb{P} that we estimate is composed of hepatic dominant phase (HADP), hepatic venous phase (HVP), native phase (NP) (without contrast injection), and equilibrium phase (EP). The algorithm uses a set of observations for contrast phase r_i obtained from the neighborhoods V_1, V_2, \dots, V_t of the image V . Each neighborhood V_j is specified by the coordinates of a bounding box within a d -dimensional image V , $V : R^d \rightarrow [0, 1]$. The set of observations

describe the intensity appearance specific for each region and phase. The goal of the phase contrast detection algorithm is to estimate for a given volume V the probability mass function $p(r_i|V)$ for each contrast phase $r_i \in \mathbb{P}$.

Estimating the phase r_i becomes difficult when the number of phases in the set \mathbb{P} is large; it is easier to distinguish between two phases than among four. Furthermore, not all observations are useful for classification of each phase, especially when the observations are the same for two or more phases (this is a case for kidney regions as can be seen from Figure 2). To address this problem, we propose a multi-level algorithm, where the number of phases $|\mathbb{P}_s|$ at each level s is smaller, $\mathbb{P}_s \subset \mathbb{P}$, and the set of observations is also smaller. In our experiments, we found that two levels achieve reliable contrast phase estimation (Figure 3). In this case, we can write

$$p(r_i|V) = \sum_k p(r_i|V, q_k)p(q_k|V), \tag{1}$$

where r_i and q_k are contrast phases estimated at level 2 and level 1, respectively.

We estimate the contrast using neighborhood image regions surrounding liver, aorta, vena cava, and kidneys. Each oriented region j is defined by the parameters $\theta_j = \{\mathbf{p}, \mathbf{r}, \mathbf{s}\}$, that specify the position (\mathbf{p}), orientation (\mathbf{r}), and size (\mathbf{s}) of the region. The set of observations inside the image neighborhood V_j is taken from the region defined by j and therefore

$$p(q_k|V) = \sum_j p(q_k|V, j)p(j). \tag{2}$$

Prior distribution $p(j)$ of regions surrounding anatomical structures is uniform. The term $p(q_k|V, j)$ specifies distribution of contrast phase q_k for region j . Furthermore, since the pose of the anatomy j is defined by parameters θ_j

$$p(q_k|V, j) = \int p(q_k, \theta_j|V)d\theta_j \tag{3}$$

$$= \int p(\theta_j|V)p(q_k|\theta_j, V)d\theta_j. \tag{4}$$

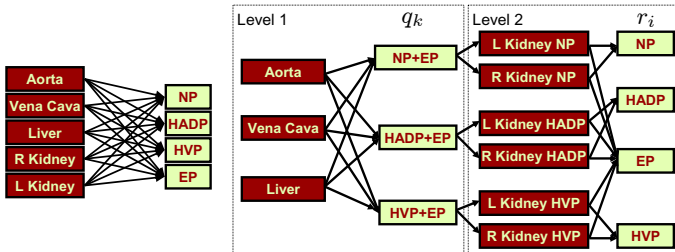


Fig. 3. Multiple level algorithm (right) has advantages compared to a single level algorithm (left) when the phase training data is imbalanced and effectively exploits discriminative power of each landmark. In our case, kidneys are useful for classifying EP phase but not the other phases (as seen in Figure 2).

In practice, we replace the integration over the parameters by the best instance

$$p(q_k|V, j) \approx p(\hat{\theta}_j|V)p(q_k|\hat{\theta}_j, V). \quad (5)$$

The set of the best instance parameters $\hat{\theta}_j$ for each anatomical region j is estimated using a volumetric context surrounding the anatomy position:

$$\hat{\theta}_j = \arg \max_{\theta_j} P(\theta_j|V), \quad (6)$$

where $P(\theta_j|V)$ is the probability of the parameters given the image volume. Let us now define a random variable $y \in \{-1, +1\}$, where $y = +1$ indicates the presence and $y = -1$ absence of the anatomy. We train a Probabilistic Boosting Tree classifier (PBT) [9] with nodes composed of AdaBoost classifiers trained to select Haar and steerable features [8] that best discriminate between positive and negative examples of the anatomy. We can then evaluate the probability of an anatomy being detected as $P(y = +1|\theta_j, V)$. A natural choice for Eq. 6 is to use the $P(y = +1|\theta_j, V)$,

$$\hat{\theta}_j = \arg \max_{\theta_j} P(y = +1|\theta_j, V). \quad (7)$$

This way, we successfully convert the above problem to a detection problem. The derivation for $p(q_k|V)$ applies to all levels of the algorithm.

We use a Multi-class LogitBoost (MLBoost) [3] classifier and a Haar feature selection from AdaBoost [3] to estimate the contrast phase distribution $p(q_k|\hat{\theta}_j, V)$ in Eq. 5. The MLBoost is a generalization of a two-class AdaBoost, interpreted using the forward additive logistic regression. The LogitBoost algorithm uses quasi-Newton steps [7] for fitting an additive symmetric logistic model by maximum-likelihood. At each iteration, the algorithm increases the classification accuracy on the training data by adding a new function $f_{km}(\hat{\theta}_j, V)$ to the response function $F_k(\hat{\theta}_j, V)$. The output of the training is a set of response functions, one for each phase q_k

$$F_k(\hat{\theta}_j, V) = \sum_m f_{km}(\hat{\theta}_j, V). \quad (8)$$

The posterior phase probability $p(q_k|\hat{\theta}_j, V)$ is then given by

$$p(q_k|\hat{\theta}_j, V) = \frac{\exp(F_k(\hat{\theta}_j, V))}{\sum_i \exp(F_i(\hat{\theta}_j, V))}. \quad (9)$$

The functions $\{f_{km}\}$ are assumed piecewise constant functions of responses computed using a set of weak classifiers selected incrementally during boosting [3].

4 Experiments

Our experiments evaluate the accuracy of individual region detectors and final classification performance.

The data set consists of 1016 CT scans with sizes ranging from $512 \times 512 \times 38$ to $512 \times 512 \times 512$ voxels and resolutions ranging from $0.52 \times 0.52 \times 0.5$ to $1.27 \times 1.27 \times 5.0$ mm resampled to a 3 mm isotropic resolution (sufficient for phase estimation application). The images were annotated by an expert, resulting in the landmark and phase annotation counts summarized in Table 1. The phase label in the Dicom tags is incorrect in 6.4% and missing in 9.1% of the cases (total of 15.5%)¹. Total of 61.8%, 29.3%, and 8.9% studies have scans from one, two, and three phases, respectively. No study contains scans from all four phases.

Table 1. Number of annotated volumes organized by each landmark and phase. The least number of volumes was obtained during the EP phase since it is not scanned as often as the other phases.

	Native	HAP	PVIP	HVP	EP	Total
Liver	81	152	128	209	17	587
Aorta	239	152	125	349	45	910
Vena Cava	174	135	104	177	42	632
Left Kidney	159	113	120	304	50	746
Right Kidney	146	158	122	314	55	795
By Phase	242	174	141	378	81	1016

In the first experiment, we assess the performance of the anatomical structure detectors. The data set with structure annotations was separated randomly into two disjoint sets, one for training (70% of volumes) and one for testing (30%). Each detector was trained using the training data set of all available contrast phases. The detectors were evaluated on the testing data set and the results compared against the annotations. The training errors of each landmark are shown in Table 2. The detection errors are low overall. Occasionally, a detector can have a larger error but this still does not mean the phase classification will be incorrect since we are using evidence from multiple anatomical regions (Eq. 5).

Table 2. Accuracy of the region detectors. The errors for vena cava are larger than aorta due to similar intensity as the liver parenchyma in NP. The vessel position and size errors along the vessel and angle rotations around the vessel are not considered due to ambiguity. The errors for kidneys and liver are higher due to their larger sizes.

	Aorta		Vena Cava		Liver		Left Kidney		Right Kidney	
	Mean	St.Dev.	Mean	St.Dev.	Mean	St.Dev.	Mean	St.Dev.	Mean	St.Dev.
Center [mm]	1.33	1.22	3.31	3.91	9.24	6.99	3.78	2.24	4.23	3.72
Angle [deg]	3.08	2.00	4.72	2.88	10.85	5.93	8.04	2.39	8.11	2.36
Size [mm]	1.01	1.30	1.00	1.72	17.81	10.60	7.04	4.40	10.05	3.28

¹ In our another database of 514 volumes, the phase label in the Dicom tags is incorrect in 5.8% and missing in 14.7% of the cases (total of 20.5%). We manually removed the language, formatting, and abbreviation ambiguities of the Dicom entries.

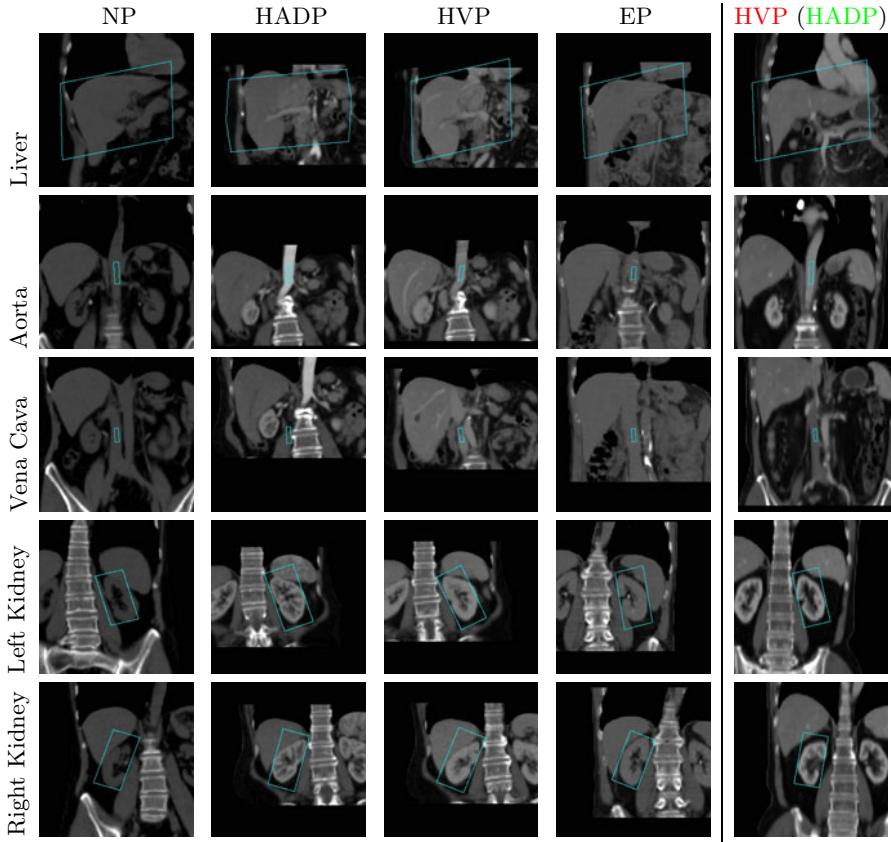


Fig. 4. Detected anatomical structures (rows) used in contrast phase estimation (cols 1-4). Anatomy enhancement specific to each phase can be clearly seen. Incorrectly classified HADP as a HVP phase for a scan in phase transition (5th col.). The lower contrast of aorta and the beginning of liver parenchyma, renal cortex and renal medulla enhancement are characteristic for a HVP phase (compare to 2nd and 3rd col.).

Finally, we present the results of the contrast phase estimation. We trained Multi-class LogitBoost [3] phase classifiers using the annotated anatomical regions. The final phase classification performance after 4-fold cross validation is summarized in a confusion Table [3]. The classification accuracy is high for NP, HADP, and HVP phases, ranging between 92.2% and 96.7%. For the EP phase, the performance is lower due to the low number of training examples. The EP phase is most often confused with HVP phase which is caused by different parts of kidney being enhanced during these phases (cortex and medulla in HVP and calices and pelvis in EP). It might be possible to improve the result by segmenting these regions to separate them for phase classifier training. The overall phase classification speed is 7 seconds on average. Example qualitative anatomical structure detection result and phase classification are shown in Figure [4].

We also compared the two-level algorithm to a single level system where each anatomical region is used to estimate all phases (Figure 3). This system has advantage compared to estimating the phase from all regions jointly since the region detection might fail. However, due to imbalanced training set (Table 1) and poor discrimination of phases in kidneys (Figure 2), the phase was correctly classified only in 85% cases with correct EP classification only in 29% cases.

Table 3. Confusion table showing percentages of correctly and incorrectly identified phases when compared to the ground truth (GT) labeling.

Detection	NP	HADP	HVP	EP
GT NP	96.2	0.8	0.00	3.0
GT HADP	2.0	92.2	5.4	0.4
GT HVP	0.5	1.1	96.7	2.2
GT EP	0.00	4.5	9.1	86.4

5 Conclusion

We presented an automatic phase classification algorithm in CT volumes. Our approach starts by a discriminative learning-based detector of anatomical structures. The regions surrounding the structures are used in Multi-class LogitBoost classifiers to accurately characterize the contrast phase. The system robustly classifies native phase (correct classification in 96.2% of the cases), hepatic dominant phase (92.2% correct), hepatic venous phase (96.7% correct), and equilibrium phase (86.4% correct). The overall speed is 7 seconds on average.

In future, we plan to exploit the time relationship between phases. This will help when there are scans from multiple phases available. We will also investigate the possibility of using pairs of anatomical regions during classification. This will limit the influence of intensity differences across regions for a particular phase as the classifier would focus on relative intensity values computed within the pair of the regions rather than on absolute intensity values in each region.

References

1. Clunie, D.: Impact of new DICOM objects on handling large data sets. In: University of Rochester PACS 2006 Conference, San Antonio, Texas (March 2006)
2. Deng, X., Du, G.: Editorial: Liver tumor segmentation. In: 3D Segmentation in the Clinic: A Grand Challenge II, MICCAI Workshop, New York, NY (2008)
3. Friedman, J., Hastie, T., Tibshirani, R.: Additive logistic regression: A statistical view of boosting. *Annals of Statistics* 28(4), 337–407 (2000)
4. Graser, A., Staehler, M., Reiser, M.F., Mueller-Lisse, U.G.: Low-dose multiphase protocols offer the best option for cancer detection, classification, and staging. *Diagnostic Imaging* (April 14, 2006)
5. Laghi, A., Sansoni, I., Celestre, M., Paolantonio, P., Passariello, R.: Computed tomography. In: Lencioni, R., Cioni, D., Bartolozzi, C. (eds.) *Focal Liver Lesions: Detection, Characterization, Ablation*. Springer, Heidelberg (2005)
6. Meyer, C., Park, H., Balter, J., Bland, P.: Method for quantifying volumetric lesion change in interval liver CT examinations. *IEEE TMI* 22(6), 776–781 (2003)
7. Nocedal, J., Wright, S.J.: *Numerical Optimization*. Springer, Heidelberg (2000)

8. Sofka, M., Zhang, J., Zhou, S., Comaniciu, D.: Multiple object detection by sequential Monte Carlo and hierarchical detection network. In: CVPR (2010)
9. Tu, Z.: Probabilistic boosting-tree: Learning discriminative models for classification, recognition, and clustering. In: CVPR, vol. 2, pp. 1589–1596 (2005)
10. Valette, P.J., Pilleul, F., Crombé-Ternamian, A.: Imaging benign and metastatic liver tumors with MDCT. In: Marchal, G., Vogl, T., Heiken, J.P. (eds.) Multidetector-Row Computed Tomography: Scanning and Contrast Protocols. Springer, Heidelberg (2005)

3D Shape Analysis for Early Diagnosis of Malignant Lung Nodules

Ayman El-Baz^{1,*}, Matthew Nitzken¹, Ahmed Elnakib¹, Fahmi Khalifa¹,
Georgy Gimel'farb², Robert Falk³, and Mohamed Abou El-Ghar⁴

¹ Bioimaging Laboratory, Bioengineering Department, University of Louisville,
Louisville, KY, USA

² Department of Computer Science, University of Auckland, Auckland, New Zealand

³ Department of Radiology, Jewish Hospital, Louisville, KY, USA

⁴ Urology and Nephrology Department, University of Mansoura, Mansoura, Egypt

Abstract. An alternative method of diagnosing malignant lung nodules by their shape, rather than conventional growth rate, is proposed. The 3D surfaces of the detected lung nodules are delineated by spherical harmonic analysis that represents a 3D surface of the lung nodule supported by the unit sphere with a linear combination of special basis functions, called Spherical Harmonics (SHs). The proposed 3D shape analysis is carried out in five steps: (i) 3D lung nodule segmentation with a deformable 3D boundary controlled by a new prior visual appearance model; (ii) 3D Delaunay triangulation to construct a 3D mesh model of the segmented lung nodule surface; (iii) mapping this model to the unit sphere; (iv) computing the SHs for the surface; and (v) determining the number of the SHs to delineate the lung nodule. We describe the lung nodule shape complexity with a new shape index, the estimated number of the SHs, and use it for the K -nearest classification into malignant and benign lung nodules. Preliminary experiments on 327 lung nodules (153 malignant and 174 benign) resulted in a classification accuracy of 93.6%, showing that the proposed method is a promising supplement to current technologies for the early diagnosis of lung cancer.

1 Introduction

A great deal of work has been published regarding the usefulness of morphologic features for discriminating malignant from benign pulmonary nodules on Computed Tomography (CT) and to a lesser extent, chest radiographs. Several studies have shown a correlation between different nodule shape characteristics and underlying pathology. For example, Furuya et al. [1] analyzed the margin characteristics of 193 pulmonary nodules on high-resolution CT and subjectively classified them as one of several types, including round, lobulated, densely spiculated, ragged, and halo. They found a high level of malignancy among the lobulated (82%), spiculated (97%), ragged (93%), and halo nodules (100%), while 66% of the round nodules proved to be benign. Automatically extracted features have

* Corresponding author:- Tel:(502)-852-5092, E-mail: aselba01@louisville.edu

also been shown to correlate with underlying malignancy. Kawata et al. [2] quantified the surface curvature and the degree of surrounding radiating pattern in biopsy-proven benign and malignant nodules, and compared the resulting feature maps. Their results showed good separation of the feature maps between the two categories. Similarly, fractal analysis has been used to quantify the nodule margin characteristics of benign and malignant nodules. Although none of these studies directly assessed the accuracy of their methods in predicting a diagnosis, they support the notion that nodule shape can potentially be used by automated systems to distinguish benign from malignant nodules. In summary, the existing shape-based approaches show the following limitations: (i) most of them classify the lung nodules based on extracted 2D features (e.g., round, lobulated, ragged, and halo, etc.) and they did not take into account the 3D features of lung nodules; (ii) most of them did not provide a quantitative measure that has the ability to describe the shape complexity of detected lung nodules; and (iii) most of the existing features (e.g., curvature, round, etc.) depend on the accuracy of the used nodule segmentation algorithm which make this process difficult for clinical practitioners to use. This work aims to address these limitations in a way that will make evaluating small lung masses more consistent.

2 Methods

2.1 Lung Nodules Segmentation

Accurate lung nodule segmentations from 3D Low Dose Computed Tomography (LDCT) images are a challenging problem because the intensities of the lung nodules and their surrounding tissues (e.g., blood vessels, etc.) are not clearly distinguishable. To overcome this problem, we use a conventional 3D parametric deformable boundary [3] and control its evolution with a new prior probabilistic visual appearance model. The prior is a 3D Markov-Gibbs Random Field (MGRF) model of the lung nodule intensities with translation- and rotation-invariant pairwise voxel interaction.

Let (x, y, z) be Cartesian 3D point coordinates. A parametric deformable surface, $\mathbf{B}(\mathbf{P}_1, \dots, \mathbf{P}_K)$, specified by K control vertices, $\mathbf{P}_k = (x_k, y_k, z_k)$, evolves in the directions that minimize its energy, E , depending on internal, $\zeta_{\text{int}}(\mathbf{B})$, and external, $\zeta_{\text{ext}}(\mathbf{B})$, forces [3]:

$$E = E_{\text{int}} + E_{\text{ext}} \equiv \int_{\mathbf{B}} (\zeta_{\text{int}}(\mathbf{B}) + \zeta_{\text{ext}}(\mathbf{B})) d\mathbf{B} \quad (1)$$

In this paper, we introduce a new type of external energy that depends on the learned prior appearance model. Let $\mathbf{Q} = \{0, 1, \dots, Q - 1\}$ and $\mathbf{L} = \{\text{nl}, \text{bg}\}$ be finite sets of image intensities (gray values) and region labels, respectively. Let a finite 3D arithmetic lattice $\mathbf{R} = [(x, y, z) : x = 0, \dots, X - 1; y = 0, \dots, Y - 1, z = 1, \dots, Z - 1]$ support a 3D image $g : \mathbf{R} \rightarrow \mathbf{Q}$ and its region map $m : \mathbf{R} \rightarrow \mathbf{L}$. The label, $m_{x,y,z}$, associates the voxel, $g_{x,y,z}$, with the lung nodule or the background. To reduce the impacts of global contrast and offset deviations of intensities due

to different sensors, each input 3D image is normalized by mapping its signal range $[q_{\min}, q_{\max}]$ to the maximal range of $[0, 255]$.

To consider the normalized images as samples of a prior MGRF model but exclude any image alignment before the segmentation, we use a generic translation- and rotation-invariant MGRF with only voxel-wise and central-symmetric pairwise voxel interaction. The latter is specified by a set \mathbf{N} of characteristic central-symmetric voxel neighborhoods $\{\mathbf{n}_\nu : \nu \in \mathbf{N}\}$ on \mathbf{R} and a corresponding set \mathbf{V} of Gibbs potentials, one per neighborhood. A central-symmetric neighborhood \mathbf{n}_ν embraces all voxel pairs such that the (x, y, z) -coordinate offsets between any voxel (x, y, z) and its neighbor (x', y', z') belong to an indexed semi-open interval $[d_{\nu, \min}, d_{\nu, \max})$; $\nu \in \mathbf{N} \subset \{1, 2, 3, \dots\}$ of the inter-voxel distances: $d_{\nu, \min} \leq \sqrt{(x - x')^2 + (y - y')^2 + (z - z')^2} < d_{\nu, \max}$.

Learning the appearance prior. Let $\mathbf{S} = \{(g_t, m_t) : t = 1, \dots, T\}$ be a training set of 3D images with known region maps. Let $\mathbf{R}_t = \{(x, y, z) : (x, y, z) \in \mathbf{R} \wedge m_t; x, y, z = \text{nl}\}$ denote the part of \mathbf{R} supporting lung nodule in the t -th training pair (g_t, m_t) ; $t = 1, \dots, T$. Let $\mathbf{C}_{\nu, t}$ be a family of voxel pairs in \mathbf{R}_t^2 with the co-ordinate offset $(\xi, \eta, \gamma) \in \mathbf{n}_\nu$ in a particular neighborhood. Let $\mathbf{F}_{\text{vx}, t}$ and $\mathbf{F}_{\nu, t}$ be an empirical marginal probability distribution of voxel intensities and of intensity co-occurrences, respectively, in the training lung nodule from \mathbf{g}_t : $\mathbf{F}_{\text{vx}, t} = [f_{\text{vx}, t}(q) = \frac{|\mathbf{R}_{t, q}|}{|\mathbf{R}_t|} : q \in \mathbf{Q}]$ and $\mathbf{F}_{\nu, t} = [f_{\nu, t}(q, q') = \frac{|\mathbf{C}_{\nu, t; q, q'}|}{|\mathbf{C}_{\nu, t}|} : (q, q') \in \mathbf{Q}^2]$ where $\mathbf{R}_{t, q} = \{(x, y, z) : (x, y, z) \in \mathbf{R}_t \wedge g_{x, y, z} = q\}$ is a subset of voxels supporting the intensity q and $\mathbf{C}_{\nu, t; q, q'}$ is a subset of the voxel pairs $\mathbf{c}_{\xi, \eta, \gamma}(x, y, z) = ((x, y, z), (x + \xi, y + \eta, z + \gamma)) \in \mathbf{R}_t^2$ supporting the intensity co-occurrence (q, q') in the training lung nodule from \mathbf{g}_t . Let $\mathbf{V}_{\text{vx}} = [V_{\text{vx}}(q) : q \in \mathbf{Q}]$ be a potential function of voxel intensities that describes the voxel-wise interaction. Let $\mathbf{V}_\nu = [V_\nu(q, q') : (q, q') \in \mathbf{Q}^2]$ be a potential function of intensity co-occurrences in the neighboring voxel pairs that describes the pairwise interaction in the neighborhood \mathbf{n}_ν ; $\nu \in \mathbf{N}$. The MGRF model of the t -th training pair is specified by the joint Gibbs probability distribution on the sublattice \mathbf{R}_t :

$$P_t = \frac{1}{Z_t} \exp(|\mathbf{R}_t| (\mathbf{V}_{\text{vx}}^\top \mathbf{F}_{\text{vx}, t} + \sum_{\nu \in \mathbf{N}} \rho_{\nu, t} \mathbf{V}_{\nu, t}^\top \mathbf{F}_{\nu, t})) \quad (2)$$

where $\rho_{\nu, t} = |\mathbf{C}_{\nu, t}|/|\mathbf{R}_t|$ is the average cardinality of \mathbf{n}_ν with respect to \mathbf{R}_t .

To identify the MGRF model in Eq. (2), the Gibbs potentials are approximated analytically¹:

$$V_{\text{vx}, \text{nl}}(q) = \log f_{\text{vx}, \text{nl}}(q) - \frac{1}{Q} \sum_{\kappa \in \mathbf{Q}} \log f_{\text{vx}, \text{nl}}(\kappa) \quad \text{for } q \in \mathbf{Q}; \text{ and} \quad (3)$$

$$V_{\nu, \text{nl}}(q, q') = \lambda \rho_\nu (f_{\nu, \text{nl}}(q, q') - f_{\text{vx}, \text{nl}}(q) f_{\text{vx}, \text{nl}}(q')) \quad \text{for } (q, q') \in \mathbf{Q}^2 \quad (4)$$

where the common factor λ is also computed analytically.

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

Boundary evolution under the appearance models. To guide the boundary evolution, we embed in the external energy term of Eq. (II) the learned prior appearance model of the detected lung nodule as follows:

$$\zeta_{\text{ext}}(\mathbf{P} = (x, y, z)) = -\pi_{\mathbf{P}}(g_{x,y,z}|\mathbf{S}) \quad (5)$$

Here, $\pi_{\mathbf{P}}(q|\mathbf{S})$ is the prior conditional probability of q , given the fixed current intensities in the characteristic central-symmetric neighborhood of \mathbf{P} for the MGRF model of Eq. (2):

$$\pi_{\mathbf{P}}(g_{x,y,z}|\mathbf{S}) = \exp(E_{\mathbf{P}}(g_{x,y,z}|\mathbf{S})) / \sum_{q \in \mathbf{Q}} \exp(E_{\mathbf{P}}(q|\mathbf{S}))$$

where $E_{\mathbf{P}}(q|\mathbf{S})$ is the conditional Gibbs energy of pairwise interaction for the voxel \mathbf{P} provided that an intensity q is assigned to the lung nodule while the other current intensities in all its neighboring voxels over the characteristic neighborhoods \mathbf{n}_{ν} ; $\nu \in \mathbf{N}$, remains fixed:

$$E_{\mathbf{P}}(q|\mathbf{S}) = V_{\text{vx},\text{nl}}(q) + \sum_{\nu \in \mathbf{N}} \sum_{(\xi,\eta,\gamma) \in \mathbf{n}_{\nu}} (V_{\nu,\text{nl}}(g_{x-\xi,y-\eta,z-\gamma}, q) + V_{\nu,\text{nl}}(q, g_{x+\xi,y+\eta,z+\gamma}))$$

After changing the energy $E_{\mathbf{B}}$ of the 3D region $\mathbf{R}_{\mathbf{B}} \subset \mathbf{R}$ inside the evolving boundary \mathbf{B} :

$$E_{\mathbf{B}} = \sum_{\forall \mathbf{P}=(x,y,z) \in \mathbf{R}_{\mathbf{B}}} E_{\mathbf{P}}(g_{x,y,z}|\mathbf{S}) \quad (6)$$

stops, the evolution terminates.

2.2 Spherical Harmonics (SHs) Shape Analysis

Spectral SH analysis [4,5] considers 3D surface data as a linear combination of specific basis functions. In our case, the surface of the segmented lung nodule is first approximated by a triangulated 3D mesh (see Fig. I) built with an algorithm by Fang and Boas [6]. Secondly, the lung nodule surface for each subject is mapped for the SH decomposition to the unit sphere. We propose a novel mapping approach, called “*Attraction-Repulsion*,” that calls for all the mesh nodes to meet two conditions: (i) the unit distance of each node from the lung nodule center, and (ii) an equal distance of each node from all of its nearest neighbors.

To detail our Attraction-Repulsion algorithm (see its summary in Algorithm 1), let τ denote the iteration index, I be the total number of the mesh nodes (in all the experiments below $I = 4896$ nodes), and $\mathbf{P}_{\tau,i}$ be the Cartesian coordinates of the surface node i at iteration τ ; $i = 1, \dots, I$. Let J be the number of the neighbors for a mesh node and $d_{\tau,ij}$ denote the Euclidean distance between the surface nodes i and j at iteration τ , where $i = 1, \dots, I$ and $j = 1, \dots, J$. Let $\Delta_{\tau,ji} = \mathbf{P}_{\tau,j} - \mathbf{P}_{\tau,i}$ denote the displacement between the nodes j and i at iteration τ . Let $C_{A,1}$, $C_{A,2}$, and C_R be the attraction and repulsion constants, respectively, that control the displacement of each surface node.



Fig. 1. Generating a 3D mesh for the lung nodule surface from a stack of successive segmented 2D LDCT slices

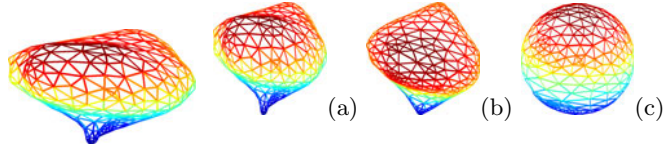


Fig. 2. Lung nodule mesh (a), its smoothed version (b), and the Attraction-Repulsion mapping to the unit sphere (c)

The starting attraction step of the proposed mapping tends to center each node, $\mathbf{P}_i; i = 1, \dots, I$, with respect to its neighbors by iteratively adjusting its location:

$$\mathbf{P}'_{\tau,i} = \mathbf{P}_{\tau,i} + C_{A,1} \sum_{j=1; j \neq i}^J \Delta_{\tau,ji} d_{\tau,ji}^2 + C_{A,2} \frac{\Delta_{\tau,ji}}{d_{\tau,ji}} \quad (7)$$

where the factor $C_{A,2}$ keeps the tightly packed nodes from collision and pushes the adjusted nodes away from their neighbors if a certain neighbor is much closer than the others.

The subsequent repulsion step inflates the whole mesh by pushing all the nodes outwards to become evenly spaced after their final back-projection onto the unit sphere along the rays from the center of the sphere. To ensure that the nodes that have not been shifted will not collide with the altered node, the location of each node, $\mathbf{P}_i; i = 1, \dots, I$, is updated before the back-projection as:

$$\mathbf{P}^{\circ}_{\tau+1,i} = \mathbf{P}'_{\tau,i} + \frac{C_R}{2I} \sum_{j=1; j \neq i}^I \left(\frac{\Delta_{\tau,ji}}{|\Delta_{\tau,ji}|^2} \right) \quad (8)$$

where a repulsion constant C_R controls the displacement of each surface node and establishes a balance between the processing time and accuracy (e.g., a smaller C_R values guarantees that the node faces will not become crossed during the iterations at the expense of the increased processing time). All the experiments below are obtained with $0.3 \leq C_R \leq 0.7$.

The original lung nodule mapped to the unit sphere with the proposed Attraction-Repulsion algorithm is approximated by a linear combination of SHs, the lower-order harmonics being sufficient to represent more generic information, while the finer details requiring the higher-order ones. The SHs are generated by solving an isotropic heat equation for the nodule surface on the unit sphere. Let $\mathcal{S} : \mathbf{M} \rightarrow \mathbf{U}$ denote the mapping of a nodule mesh \mathbf{M} to the unit sphere \mathbf{U} . Each node $\mathbf{P} = (x, y, z) \in \mathbf{M}$ mapped to the spherical position $\mathbf{u} = S(\mathbf{P})$ is represented by the spherical coordinates $\mathbf{u} = (\sin \theta \cos \varphi, \sin \theta \sin \varphi, \cos \theta)$ where $\theta \in [0, \pi]$ and $\varphi \in [0, 2\pi)$ are the polar and azimuth angles, respectively. The SH $Y_{\alpha\beta}$ of degree α and order β is defined as [7]:

Algorithm 1: Attraction-Repulsion Algorithm

Initialization

- Construct the 3D lung nodule mesh (Fig. 2,a).
- Smooth it by the Laplacian filtering (Fig. 2,b).
- Initialize the mapping of the smoothed mesh to the unit sphere.

Repeat

- **For** $i = 1 \rightarrow I$
 - **Attraction:**
 - * Select a node to process.
 - * Update the node using Eq. (7)
 - **Repulsion:**
 - * Update the node using Eq. (8).
- **End** (all nodes in the mesh are shifted and back-projected onto the unit sphere).

While changes in the node positions occur (Fig. 2,c).













	B	M
OM		
1 SH		
5 SHs		
10 SHs		
15 SHs		
60 SHs		

Fig. 3. Approximation of the 3D shape for malignant (M), benign nodules (B), and original mesh (OM)

$$Y_{\alpha\beta} = \begin{cases} c_{\alpha\beta} G_{\alpha}^{|\beta|} \cos \theta \sin(|\beta|\varphi) & -\alpha \leq \beta \leq -1 \\ \frac{c_{\alpha\beta}}{\sqrt{2}} G_{\alpha}^{|\beta|} \cos \theta & \beta = 0 \\ c_{\alpha\beta} G_{\alpha}^{|\beta|} \cos \theta \cos(|\beta|\varphi) & 1 \leq \beta \leq \alpha \end{cases} \quad (9)$$

where $c_{\alpha\beta} = \left(\frac{2\alpha+1}{2\pi} \frac{(\alpha-|\beta|)!}{(\alpha+|\beta|)!}\right)^{\frac{1}{2}}$ and $G_{\alpha}^{|\beta|}$ is the associated Legendre polynomial of degree α and order β . For a fixed α , the polynomials G_{α}^{β} are orthogonal over the range $[-1, 1]$. As shown in [7], the Legendre polynomials are effective in calculating SHs. This is the main motivation behind their use in this work.

Finally, the lung nodule is reconstructed from the SHs of Eq. (9). In the case of the SHs expansion, the standard least-square fitting does not accurately model the 3D shape of the lung nodule and can miss some of the shape details that discriminate between the malignant and benign lung nodules. To circumvent this problem, we used the iterative residual fitting by Shen et al. [8] that accurately approximates the 3D shapes of malignant and benign lung nodules. As shown in Fig. 3, the model accuracy does not significantly change for the benign nodule from 15 to 60 SHs, while it continues to increase for the malignant nodule.

2.3 Quantitative Lung Nodule Shape Analysis

Our main hypothesis is that the shape of malignant nodules is more complicated (e.g., with spiculation) when compared with the shape of benign nodules which are simpler (smoothed shape) as in Fig. 3, so that more SHs have to be used for

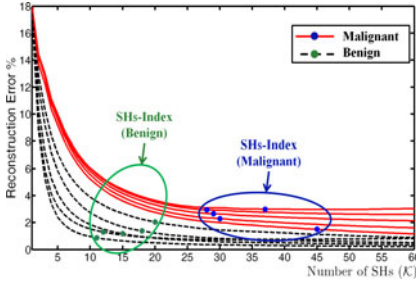


Fig. 4. Estimation of the shape index from the total nodule approximation error for malignant and benign nodules

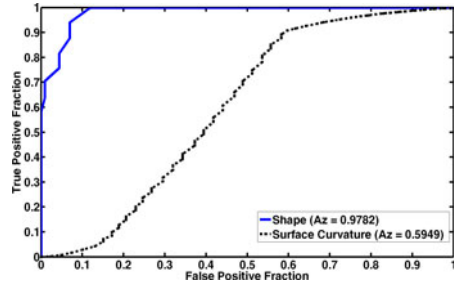


Fig. 5. The ROC curves for our approach and the surface curvature-based diagnostic approach

accurate approximation of the shape of malignant lung nodule. Therefore, the number of the SHs after which there are no significant changes in the approximations can be used as a new shape index that quantifies the shape complexity of the detected lung nodules. Due to the unit sphere mapping, the original mesh for each nodule is inherently aligned with its reconstructed mesh shape, and the sum of the Euclidean distances between the corresponding nodes gives the total error between both the mesh models. As shown in Fig. 4, the total error curves for the increasing number \mathcal{K} of the SHs can be statistically analyzed to differentiate between the detected lung nodules.

3 Experimental Results and Conclusions

To justify the proposed methodology of analyzing the 3D shape of both malignant and benign nodules, the above proposed shape analysis framework was pilot-tested on a database of clinical multislice chest LDCT scans of 327 lung nodules (153 malignant and 174 benign). The CT data sets each have $0.7 \times 0.7 \times 2.0 \text{ mm}^3$ voxels, with nodule diameters ranging from 3 mm to 30 mm. Note that these 327 nodules were diagnosed using either bronchoscopy and needle biopsy, or two-year follow-up with CT scans². Also, our current database does not contain Ground Glass Nodules (GGN).

The training subset for classification (15 malignant lung nodules and 15 benign lung nodules) were arbitrarily selected from 327 lung nodules. The accuracy of classification based on using a K-nearest classifier for both the training and test subjects was evaluated using the χ^2 -test at 95% confidence level. At the 95% confidence level, 143 out of 153 malignant nodules (a 93.5% accuracy) were correctly classified, and 163 out of 174 control subjects (a 93.7% accuracy) were correctly classified. The overall accuracy using the proposed 3D shape-based CAD system is 93.6% in the first detection of lung nodules. The classification based on the

² For complete details about our nodules database, please see: <https://louisville.edu/speed/bioengineering/faculty/bioengineering-full/dr-ayman-el-baz/supplemental-materials>.

traditional analysis of surface curvature-based diagnostic approach [2] correctly classifies 77 out of 153 malignant nodules (a 50.3% accuracy), and 103 out of 174 benign nodules (a 59.2% accuracy) at a 95% confidence level. These results highlight the advantage of the proposed approach.

Another way to measure and test the performance of the proposed diagnostic system is to compute the Receiver Operating Characteristic (ROC). Each point on the graph is generated by using a different cut point (classification threshold). Figure 5 shows the ROC curves of the two approaches, our proposed shape index-based diagnostic approach and the surface curvature-based diagnostic approach [2]. It is clear from the data in Fig. 5 that the area under the ROC curve of our present approach is much larger ($Az = 0.9782$) than the area under the ROC curve of the surface curvature-based diagnostic approach [2] ($Az = 0.5949$). The high sensitivity and specificity of the proposed approach is due to using the estimated number of spherical harmonics to approximate the 3D shape of a detected lung nodule as a new discriminatory feature which is more separable than using surface curvature. More experimental results that address the sensitivity of our approach w.r.t. the accuracy of segmentation and mesh generation steps has been posted on our web site².

As demonstrated in this paper, the preliminary results justify the elaboration of the proposed alternative method for diagnosing malignant lung nodules. Its novelty lies in using the shape of a 3D nodule instead of the more conventional surface curvature as a reliable diagnostic feature. The shape is described in terms of a linear combination of SHs.

References

1. Furuya, K., Murayama, S., Soeda, H., Murakami, J., Ichinose, Y., Yabuuchi, H., Katsuda, Y., Koga, M., Masuda, K.: New Classification of Small Pulmonary Nodules by Margin Characteristics on High-resolution CT. *Acta Radiol.* 40(5), 496–504 (1999)
2. Kawata, Y., Niki, N., Ohmatsu, H., Kakinuma, R., Eguchi, K., Kaneko, R., Moriyama, N.: Quantitative Surface Characterization of Pulmonary Nodules based on Thin-section CT Images. *IEEE Trans. Nucl. Sci.* 45(4), 2132–2138 (1998)
3. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active Contour Models. *Int. J. Comput. Vision* 1(4), 321–331 (1987)
4. Chung, M.K., Evans, A.C., Davidson, R.J.: Weighted Fourier Series Representation and its Application to Quantifying the Amount of Gray Matter. *IEEE Trans. Med. Imag.* 26(4), 566–581 (2007)
5. Nain, D., Haker, S., Bobick, A., Tannenbaum, A.: Multiscale 3D Shape Analysis using Spherical Wavelets. In: Duncan, J.S., Gerig, G. (eds.) *MICCAI 2005*. LNCS, vol. 3750, pp. 459–467. Springer, Heidelberg (2005)
6. Fang, Q., Boas, D.: Tetrahedral Mesh Generation from Volumetric Binary and Grayscale Images. In: 6th IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI 2009), pp. 1142–1145 (2009)
7. Courant, R., Hilbert, D.: *Methods of Mathematical Physics*. John Willy&Sons, New York (1989)
8. Shen, L., Chung, M.K.: Large-scale Modeling of Parametric Surfaces using Spherical Harmonics. In: 3rd International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT 2006), pp. 294–301 (2006)

Classification of Diffuse Lung Disease Patterns on High-Resolution Computed Tomography by a Bag of Words Approach

Rui Xu¹, Yasushi Hirano¹, Rie Tachibana², and Shoji Kido¹

¹ Applied Medical Engineering Science, Graduate School of Medicine,
Yamaguchi University, Ube, Japan

{xurui,yhirano,kido.ai}@yamaguchi-u.ac.jp

² Information Science and Technology Dept.,
Oshima National College of Maritime Technology, Oshima-Gun, Japan
tatibana@oshima-k.ac.jp

Abstract. Visual inspection of diffuse lung disease (DLD) patterns on high-resolution computed tomography (HRCT) is difficult because of their high complexity. We proposed a bag of words based method on the classification of these textural patterns in order to improve the detection and diagnosis of DLD for radiologists. Six kinds of typical pulmonary patterns were considered in this work. They were consolidation, ground-glass opacity, honeycombing, emphysema, nodular and normal tissue. Because they were characterized by both CT values and shapes, we proposed a set of statistical measure based local features calculated from both CT values and the eigen-values of Hessian matrices. The proposed method could achieve the recognition rate of 95.85%, which was higher comparing with one global feature based method and two other CT values based bag of words methods.

1 Introduction

Diffuse lung disease (DLD) refers to a group of lung disorders which spread out in large areas. In the detection and diagnosis of DLD, high-resolution computed tomography (HRCT) has played important roles in recent years [12]. Thin slice CT is able to give detailed appearances on pulmonary patterns characterized for specific abnormal findings which point toward a specific diagnosis and treatment. In current clinical protocols, the objective identification of such patterns has not been established, and visual inspection is carried on, according to the experiences of radiologists. Not only subjective differences lead to inevitable misjudgements, but also huge amount of images makes a big burden on radiologists. Therefore, a quantitative computer-aided diagnosis (CAD) tool is required to give the second opinion to facilitate the detection and diagnosis of DLD.

There were researches on computer-aided analysis of DLD patterns in the past ten years [9]. From the viewpoints of computer vision, it can be generalized as the problem of texture analysis on a certain 2D or 3D region of interest (ROI). Some classical textural feature analysis methods calculated on 2D ROIs

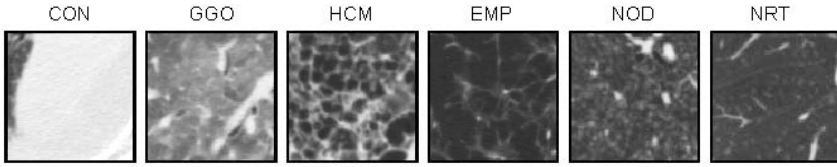


Fig. 1. Examples of six kinds of pulmonary patterns in HRCT

can be used, including the direct analysis on gray value intensities [2], features calculated on histogram statistics, and features extraction based on filter-banks, co-occurrence matrices or run-length parameters [1] [5] [13]. Features may also be designed for the specific task on hands [4]. Different from the above-mentioned methods which analyzed on 2D ROIs, a method based on features calculated on 3D ROIs was proposed in [10], which reported that 3D ROIs based features were more sensitive and specific. All of these textural analysis methods made use of global features.

As the development of computer vision techniques, there have been more and more researches which reported that techniques based on local features could lead to more accurate results on recognition tasks [11]. Local features are less sensitive to the changing of illuminations and positions, so they usually could be a more robust way to represent images of underlying objects. An efficient approach to use local features is called bag of words (keypoints) [6]. It was firstly proposed in the area of statistical natural language analysis [3], then introduced into the area of computer vision in [6] for many applications.

Recently, local features have been considered on the analysis of pulmonary patterns on HRCT. Local binary patterns (LBP) combined with k nearest neighbor (k -NN) was adopted to distinguish three kinds of pulmonary patterns (normal tissues, centrilobular emphysema and paraseptal emphysema) for chronic obstructive pulmonary disease (COPD) in [16]. A bag of words approach (originally called textron-based method) was also tried in the same problem in [15]. According to our knowledge, there are no works to apply local features on DLD patterns analysis. Compared to [15], the main differences are listed as follows:

- our aim is to classify pulmonary patterns for DLD rather than COPD.
- more categories of pulmonary patterns.
- local features extracted from 3D ROIs instead of 2D ROIs.
- compact and efficient statistical measure based local features considering both shapes and CT values.

Fig. 1 gives the examples of six types of typical pulmonary patterns in HRCT for DLD. They are consolidation (CON), ground-glass opacity (GGO), honeycombing (HCM), emphysema (EMP), nodular (NOD) and normal tissue (NRT). The aim of our work is to find an efficient computer-aided analysis method to distinguish them in order to improve the detection and diagnosis of DLD on HRCT.

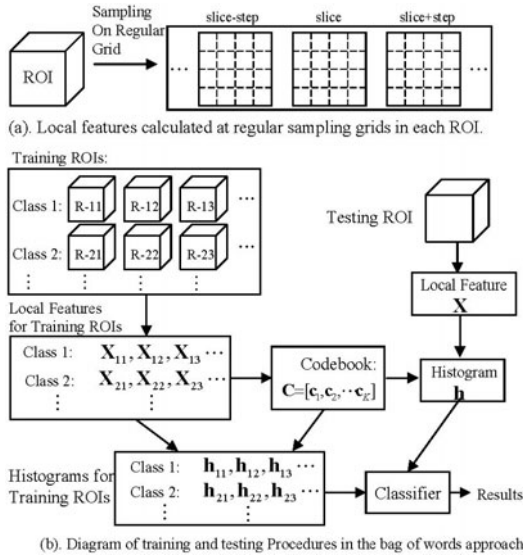


Fig. 2. The framework of the bag of words approach

2 Method

2.1 Bag of Words Approach

Fig. 2 describes the framework of bag of words approach, which is composed by three main steps, local feature calculation, code-book construction and classification. Local features can be calculated at the regularly sampling grids or at the pre-determined key-point positions. According to literature [11], there seems to be no evidence to show which one is better. In this work, we prefer to adopt the way of regularly sampling grids, since it is easier and faster. Here, the ROI size and the grid step is set to be $32 \times 32 \times 16$ and $4 \times 4 \times 4$ respectively, so we can get $9 \times 9 \times 5 = 405$ sampling points on one ROI. For each sampling point, one local feature is calculated in a patch whose center is located on it. Therefore, 405 local features can be calculated on each ROI. The patch size is the parameter determined by experiments.

The main idea of bag of words approach is to represent the ROIs (or images) by a histogram whose bins are the elements of a code-book trained from local features. Such a code-book can be seen as an intermediate layer to interpret images. Since local features are calculated from limited local regions, each of them only reflects information of partial objects. Although local features are huge, they are usually clustered into limited number of centers in high-dimensional feature spaces. Just as words are basic elements of sentences, such centers of clusters can also be considered as the basic elements which compose the underlying objects. This is the reason why it is called bag of words approach. The unit of these elements is called a code-book or dictionary. Construction of such a

dictionary can be generalized as a vector quantization problem. K-means method is adopted to resolve this problem. The number of clusters in K-means (K) is another parameter determined by experiments.

In the classification step, all of the local features of one ROI are assigned to a label which indicates the nearest elements (or words) in the code-book. Given each elements as the bins, a histogram can be calculated to refer to the statistical information of the ROI. Since the same type of ROIs gives similar histograms, such histograms can be treated as input vectors for the classification.

In our opinion, bag of words approach is a common framework. According to different recognition tasks, the method of local feature extraction and classifier should be adjusted in order to make it achieve better results. We will dwell on these two aspects in the following two subsections.

2.2 Local Features

CT values of a squared patch was treated as local features for each 2D ROIs in [15]. It was reported that such local features achieved good results for the recognition of normal tissues and two kinds of emphysema patterns. We tried this method, but the results were not satisfied. Therefore, we reconsidered about our problem and designed a new kind of local feature extraction method suitable for our task.

The categories of pulmonary patterns in our work are more than [15]. Some patterns, such as NRT, CON and EMP, are mainly characterized by CT values. However, shape information should be taken into consideration in order to distinguish the patterns, such as HCM and NOD. Since shape information can not be shown from 2D slices, 3D ROIs are adopted in this work. An usual way to describe shape information is to use the eigen-values of Hessian matrices calculated on 3D ROIs. These eigen-values can be arranged to form three cubes according to the arrangement of the original voxels. In order to consider the information of both CT values and shapes, local features are calculated from both the three eigen-value cubes and the original ROI. Four kinds of statistical measures, mean, variance, skewness and kurtosis are calculated from a cube-patch centered at the regular sampling grids. Therefore, one local features consists of 16 statistical measures. The size of cube-patch is the parameter for this local feature extraction method and its optimal value is determined by experiments. It should be noted that no matter how it changes, the compactness of such local features does not change. Additionally, these statistical measures are invariant to translation and rotation. Here, the Hessian matrix was only calculated on the pixel level.

2.3 Classifier

SVM was adopted as the classifier in this work. We used a version called LIB-SVM [17]. For a generalized recognition problem, a common choice of the kernel for SVM is a Gaussian kernel. For the bag of words problem, it was reported that some other kernels was superior to Gaussian kernel if input vectors were

histograms [14]. We tried other three kinds of kernels. The kernel definitions are given by Eq. 1

$$\begin{cases} G(\mathbf{h}, \mathbf{h}') = \exp(-\gamma \|\mathbf{h} - \mathbf{h}'\|^2) \\ K_1(\mathbf{h}, \mathbf{h}') = \sum_{i=1}^N \min(h_i, h'_i) \\ K_2(\mathbf{h}, \mathbf{h}') = \sum_{i=1}^N \sqrt{h_i h'_i} \\ K_3(\mathbf{h}, \mathbf{h}') = \exp(-\alpha \sum_{i=1}^N \frac{(h_i - h'_i)^2}{h_i + h'_i}) \end{cases} \quad (1)$$

where $\mathbf{h} = [h_1, \dots, h_N]^T$ is the N -bin histogram. $G(\mathbf{h}, \mathbf{h}')$ is the *Gaussian* kernel, $K_1(\mathbf{h}, \mathbf{h}')$ is the *histogram intersection* kernel, $K_2(\mathbf{h}, \mathbf{h}')$ is the *Bhattacharyya* kernel, $K_3(\mathbf{h}, \mathbf{h}')$ is the χ^2 -kernel. γ is the parameter for the gaussian kernel and α is the parameter for the χ^2 -kernel. Additionally, a soft margin parameter C is considered for all the four kernels. These parameters will be optimized in experiments.

3 Experiments and Results

3.1 Data

We obtained 117 scans for different subjects from Tokushima University Hospital, Tokushima, Japan. All of them were scanned from 16-row multi-slice CT (Aquilion, Toshiba Co.), when edge-enhanced filtering was not applied. The resolution was 512×512 with the pixel size of 0.6mm on each slice, and the slice-thickness was 1mm. The regions of the six types of patterns were marked by three experienced radiologists according to the following procedure. Firstly, one radiologist was asked to review all scans. From each scan, maximum of three slices were selected where typical patters dominantly spread. Then together with the other two radiologists, the six kinds of patters were marked on the selected slices separately. Finally, the common regions marked by all radiologists were considered as where typical patterns were located. The ROIs were constructed according to these determined regions. The centers of ROIs were randomly selected from them while considering non-overlap on ROIs. At last, 3009 3D-ROIs were determined.

3.2 Results and Discussion

There were mainly three kinds of parameters which should be adjusted by experiments in our proposed method. They were the size of cube-patches, the number of clusters in K-means, and the parameters related to SVM classifiers (including kernel types, the soft margin parameter, and parameters for each kernel). These parameters were determined by a 20-fold cross-validation test. First, by setting number of clusters to be a certain number (100 was used) and the kernel type of SVM to be the Gaussian kernel, we exhausted the possible values on the size of cube-patches, soft margin parameter and γ of the gaussian kernel to train classifiers. The size of cube-patches was optimized by testing on the training data of 20-fold cross-validation tests. Second, by using the optimized cube-patch

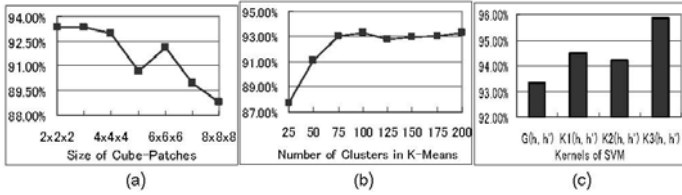


Fig. 3. The choices of parameters in our method

size and setting the classifier to be Gaussian kernel based SVM again, we tried all the possible values on the number of clusters and parameters related to the classifier to optimize the number of clusters in the same way. Finally, the parameters related to SVM were optimized, when the cube-patch size and the number of clusters were set to be the optimized values.

For the size of the cube-patches, we tried the size from $2 \times 2 \times 2$ to $8 \times 8 \times 8$. Fig. 3 (a) gives the results when different patch sizes were used. Considering that more voxels could give more stable statistical measures, the size of $3 \times 3 \times 3$ was determined in this work. On the choice of the number of clusters in K-Means, we tried the number from 25 to 200 (Fig. 3 (b)) and chose 100 as the optimized value. According to the experimental results, shown by Fig. 3 (c), the χ^2 -kernel ($K_3(\mathbf{h}, \mathbf{h}')$) gave a little better results than the other three kernels. We chose the χ^2 -kernel as the optimal kernel for the SVM classifier.

We also compared our method with one global feature based method (Glo-3D) and two bag of words methods (CTV-2D and CTV-3D) which only made use of the CT values. In order to fairly compare them with our method, all their parameters were optimized by the same way used for our method. A brief introduction about them is listed as follows:

- **CT Values on 2D Patches (CTV-2D).** We implemented a similar version of the method proposed in [15]. The ROI was a 2D slice whose center was the same as the 3D-ROI used in this work. The CT values in 2D squared-patches were arranged to be a vector as local features. According to [15], SVM with Gaussian kernels was adopted as the classifier. The parameters are the size of squared-patches, the number of clusters, the soft margin parameter and the γ of Gaussian kernel.
- **CT Values on 3D Patches (CTV-3D).** This method was similar to CTV-2D. The difference was that a 3D ROI was used instead of a 2D ROI. CT values based local features and Gaussian kernel based SVM were used. The parameters are the size of cube-patches, the number of clusters, the soft margin parameter and the γ of Gaussian kernel.
- **Global Features on 3D Patches (Glo-3D).** This was not a bag of words approach. Global features calculated directly from a 3D ROI consisted of two kinds of common used textural features, measures on gray-level co-occurrence matrices (GLCM) [7] and measures on gray-level run-length matrices (GLRLM) [8]. The two kinds of measures were concatenated to be a

Table 1. Recognition results in experiments

(a) Comparison of four methods

Methods	Accuracy
CTV-2D	86.71%
CTV-3D	91.12%
Glo-3D	91.96%
Ours	95.85%

(b) Confusion table of our method

True Labels	Estimated Labels						Accuracy
	CON	GGO	HCM	EMP	NOD	NRT	
CON	122	1	1	0	0	0	98.39%
GGO	1	494	2	1	8	5	96.67%
HCM	0	2	520	3	1	0	98.86%
EMP	0	3	5	710	6	13	96.34%
NOD	1	9	0	11	313	26	86.94%
NRT	0	5	0	2	19	725	96.54%

feature vector for recognition. SVM with Gaussian kernels was adopted as the classifier. The parameters are the soft margin parameter and the γ of Gaussian kernel.

Using the parameters optimized in training, we compared the four methods by testing on the testing data in the 20-fold cross-validation tests. Table 1(a) gives the comparison results. It can be seen that our methods gave higher recognition accuracy than the other three methods. It should be noted that Glo-3D gave better results than the two bag of words based methods, CTV-2D and CTV-3D. This is because that the only consideration of CT values in local features can not distinguish these patterns and affects the performances of bag of words approaches. Table 1(b) gives the confusion table for each kinds of patterns for our method.

4 Conclusion

We proposed a bag of word approach to automatically classify six kinds of pulmonary patterns on HRCT for DLD. Some patterns, such as CON, NRT and EMP, were mainly characterized by CT values; while for other patterns, such as NOD and HCM, both CT values and shape information should be considered in order to classify them successfully. According to such a consideration, we proposed a new kind of local features calculated from both the original CT values and eigen-values of Hessian matrices for our bag of words approach. Experimental results showed that this method was superior to other two kinds of bag of words approaches which only depended on CT values and one global feature based method. We will try some other local features, or combine both local and global features in our future research.

Acknowledgement. This research was supported by a grand-in-aid for scientific research from the Ministry of Education, Culture, Sports, Science, and Technology, Japan, under the grand No. 21103008.

References

1. Uppaluri, R., Heitmman, E.A., Sonka, M., Hartley, P.G., Hunninghake, G.W., Mclennan, G.: Computer Recognition of Regional Lung Disease Patterns. American Journal of Respiratory and Critical Care Medicine 160(2), 648–654 (1999)

2. Kauczor, H., Heitmann, K., et al.: Automatic Detection and Quantification of Ground-glass opacities on High-resolution CT Using multiple Neural Networks: Comparison with a Density Mask. *American Journal of Roentgenology* 175(5), 1329–1334 (2000)
3. Manning, C.D., Schütze, H.: *Foundation of Statistical Natural Language Processing*. The MIT Press, Cambridge (2001)
4. Uchiyama, Y., Katsuragawa, S., Abe, H., Shiraiishi, J., Li, F., Li, Q., Zhang, C., Suzuki, K., Doi, K.: Quantitative Computerized Analysis of Diffuse Lung Disease in High-resolution Computed Tomography. *Medical Physics* 30(9), 2440–2454 (2003)
5. Sluimer, I.C., Waes, P.F.V., Viergever, M.A., Ginneken, B.V.: Computer-aided Diagnosis in High-resolution CT of the Lungs. *Medical Physics* 30(12), 3081–3090 (2003)
6. Csurka, G., Dance, C.R., Fan, L., Bray, C.: Visual Categorization with Bags of Keypoints. In: *Proc. of ECCV Workshop on Statistical Learning in Computer Vision*, pp. 1–22 (2004)
7. Kurani, A.S., Xu, D., Furst, J.D., Raicu, D.S.: Co-occurrence matrices for volumetric data. In: *7th IASTED International Conference on Computer Graphics and Imaging*, Hawai (2004)
8. Xu, D., Kurani, A.S., Furst, J.D., Raicu, D.S.: Run-length encoding for volumetric texture. In: *The 4th IASTED International Conference on Visualization, Imaging, and Image Processing* (2004)
9. Sluimer, I., Schilham, A., Prokop, M., Ginneken, B.V.: Computer Analysis of Computed Tomography Scans of the Lung: A Survey. *IEEE Transactions on Medical Imaging* 25(4), 385–405 (2006)
10. Xu, Y., Sonka, M., McLennan, G., Guo, J.F., Hoffman, E.A.: MDCT-based 3-D Texture Classification of Emphysema and Early Smoking Related Lung Pathologies. *IEEE Transactions on Medical Imaging* 25(4), 464–475 (2006)
11. Tuytelaars, T., Mikolajczyk, K.: Local Invariant Feature Detectors: A Survey. *Foundations and Trends in Computer Graphics and Vision* 3(3), 177–280 (2007)
12. Webb, W., Müller, N.L., Naidich, D.: *High Resolution CT of the Lung*, 4th edn. Lippincott Williams & Wilkins, Baltimore (2008)
13. Park, S.O., Seo, J.B., Kim, N., Park, S.H., Lee, Y.K., Park, B.W., Sung, Y.S., Lee, Y., Kang, J., Lee, J., Kang, S.H.: Feasibility of Automated Quantification of Regional Disease Patterns Depicted on High-resolution Computed Tomography in Patients with Various Diffuse Lung Diseases. *Korean Journal of Radiology* 10(5), 455–463 (2009)
14. Lampert, C.H.: Kernel Methods in Computer Vision. *Foundations and Trends in Computer Graphics and Vision* 4(3), 193–285 (2009)
15. Gangeh, M.J., Sørensen, L., Shaker, S.B., Kamel, M.S., Bruijne, M.D., Loog, M.: A Texton-Based Approach for the Classification of Lung Parenchyma in CT Images. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) *MICCAI 2010*. LNCS, vol. 6363, pp. 595–602. Springer, Heidelberg (2010)
16. Sørensen, L., Shaker, S.B., de Bruijne, M.: Quantitative analysis of pulmonary emphysema using local binary patterns. *IEEE Transaction on Medical Imaging* 29(2), 559–569 (2010)
17. Chang, C.C., Lin, C.J.: LIBSVM : a library for support vector machines, <http://www.csie.ntu.edu.tw/~cjlin/libsvm>

Discriminative Pathological Context Detection in Thoracic Images Based on Multi-level Inference

Yang Song¹, Weidong Cai¹, Stefan Eberl^{1,2}, Michael J. Fulham^{1,2,3},
and Dagan Feng^{1,4}

¹Biomedical and Multimedia Information Technology (BMIT) Research Group,
School of Information Technologies, University of Sydney, Australia

²Department of PET and Nuclear Medicine, Royal Prince Alfred Hospital,
Sydney, Australia

³Sydney Medical School, University of Sydney, Australia

⁴Center for Multimedia Signal Processing (CMSP), Department of Electronic &
Information Engineering, Hong Kong Polytechnic University, Hong Kong

Abstract. Positron emission tomography - computed tomography (PET-CT) is now accepted as the best imaging technique to accurately stage lung cancer. The consistent and accurate interpretation of PET-CT images, however, is not a trivial task. We propose a discriminative, multi-level learning and inference method to automatically detect the pathological contexts in the thoracic PET-CT images, i.e. the primary tumor and its spatial relationships within the lung and mediastinum, and disease in regional lymph nodes. The detection results can also be used as features to retrieve similar images with previous diagnosis from an imaging database as a reference set to aid physicians in PET-CT scan interpretation. Our evaluation with clinical data from lung cancer patients suggests our approach is highly accurate.

1 Introduction

Lung cancer is among the most common malignancies in the Western world, and accurate staging is critical for the selection of the most appropriate therapy, be it surgery, chemotherapy, radiotherapy or combined therapies. The size and extent of the primary tumor and the status of mediastinal lymph nodes are critical for staging the thorax; automated methods to achieve this goal can shorten the time a physician needs to read an image.

PET-CT is now accepted as the best imaging technique to accurately stage the most common form of primary lung cancer, non-small cell lung cancer (NSCLC). PET-CT scanners produce co-registered anatomical (CT) and functional (PET) patient information from a single scanning session. The PET tracer ¹⁸F-*fluoro-deoxy-glucose* (FDG) is the most commonly used tracer for clinical PET-CT diagnosis, and tumors typically take up more FDG than surrounding normal structures.

Our aim is to develop a method to automatically detect the primary tumor, the spatial relationships of the tumor within the lung and to the mediastinum, and the location of disease in lymph nodes. The objective is not to perform a precise segmentation, but to provide an inference of the pathological context and function as a robust localization system to assist the reading physician. The detection output can also serve as input to a content-based image retrieval (CBIR) system to retrieve similar imaging cases to help interpretation.

Related work. The majority of existing work focuses on segmentation on CT images using various classification techniques [1,2,3]. Our method is partially motivated by these approaches. However, they do not support concurrent detection of tumors and abnormal lymph nodes, and do not consider the complexity caused by two pathological types within one image. Recent work by Wojak et.al. introduced a tumor and lymph node segmentation method on PET-CT images using energy minimization [4]. However, the work does not address the differentiation between tumors and lymph nodes, and the spatial context of the tumors.

The work most similar to ours was reported by Wu et.al. [5], for detecting lung nodules and the connectivity with vessel, fissure and lung wall, and did not aim for perfect segmentations. However, it differs from our approach in several aspects: (1) it works on CT subvolumes with the nodule appearing at the center, while our method works on raw PET-CT images of the entire thorax; (2) our method detects abnormal lymph nodes and differentiates them from the primary tumors; and (3) we are interested in the higher-level spatial relationships, i.e. the connectivity between tumors and the chest wall and mediastinum.

Our work has also been provoked by the idea of multi-class object detection proposed for general computer vision problems [6,7,8]. Different from these methods, we design three levels of features to exploit the specific characteristics of PET-CT thoracic images, and a different multi-level discriminative model for more effective inference of the pathological context.

2 Method

2.1 Discriminative Structure Localization

At the first stage we detected four types of structures – the lung fields (L), mediastinum (M), tumor (T) and disease in lymph nodes (N) – from the thoracic images. We formulated the detection as a multi-level, multi-class (L, M, T or N) object localization problem. For an image \mathbf{I} , the classification score with labeling \mathbf{Y} (the label matrix of \mathbf{I}) is defined as:

$$S(\mathbf{I}, \mathbf{Y}) = \sum_l \alpha_{y_l} \cdot f_l + \sum_s \beta_{y_s} \cdot f_s + \sum_o \gamma_{y_o} \cdot f_o \quad (1)$$

where f_l , f_s and f_o are the three levels of feature vectors (local, spatial, and object levels) of \mathbf{I} ; α , β and γ are the respective feature weights; y_l , y_s and y_o are the class labels at each level, representing the four classes; and l , s and o are

the indices of the regions formed at each level in the transaxial slices. The goal was to find the labeling \mathbf{Y} that maximized the score S for image \mathbf{I} .

Our approach was region-based for effective modeling of the higher-level features, and we designed a cascaded learning approach for the classification. The higher-level spatial and object features were important for differentiating the four types of structures, especially for between T and N, and between T and M, as described in more details in the following sections. We also employed a two-phase design by exploring first the 2D features at the local and spatial levels, then the 3D features at the object level; this was to optimize the classification for each image slice first before considering the inter-slice relationships.

Local-level Modeling. Each image slice was first clustered into a number of regions of various sizes and shapes using the mean-shift algorithm [9]. The regions were generated separately for PET and CT slices, and then merged into one set for each slice pair. Each region R_l was then represented by the local feature f_l : the mean CT density; and the mean standardized uptake values (SUV), which was computed by normalizing the mean SUV of R_l based on an adaptive threshold [10].

At the local level, f_l could not differentiate between T and N, because both had high CT densities and high SUV values. So, we limited y_l to take three values, L, M or T/N, to focus on differentiating the pathological tissues from lung fields and mediastinum.

Spatial-level Modeling. Besides an inability to distinguish T and N, another major problem with local-level modeling was that areas surround the tumor were often misclassified as M, which could subsequently cause T to be confused as N. To better classify the surrounding area, we observed that the spatial information played an important part, e.g. its proximity with T and L and distance from M, and the differences between its average CT density and SUV and those of the other regions. Similar spatial features could also help to improve the labeling of some misclassified regions in the mediastinum.

The spatial-level features were thus computed as the following feature vector f_s for region R_s in 11 dimensions: (Dim. 1-3) the average spatial distance from region R_s to other regions R_i of type k ($k \in \{L, M, T/N\}$); (Dim. 4) the size of R_s ; (Dim. 5-7) the difference between the mean CT density of R_s and the average CT densities of all regions of type k ; (Dim. 8-10) the difference between the mean SUV of R_s and the average SUV of all regions of type k ; and (Dim. 11) the local-level labeling at R_s .

The regions R_s at this level were different from the local-level ones. We first performed another mean-shift clustering for areas around the detected abnormal regions, to discover finer-scale details. For regions not connected with the abnormal areas, and with high confidence of being L or M (based on the classification score), we also merged the connected regions of the same type into one region. And similarly to y_l , y_s could be either L, M or T/N.

Object-level Modeling. So far, T and N were still treated as one type, and the transaxial slices were processed separately. Based on the classification results

of the previous level, by merging connected regions with the same label into one region, a slice was then represented as a relatively small number of regions, roughly corresponding to the anatomical structures, but with some discontinuous segments. The goal was thus to differentiate tumors from abnormal lymph nodes and smooth the labeling, and we observed that the object-level information was the main distinctive factor. For example, T should be within L and possibly invading into M while N should be within M; hence, the distance between T and L regions should be small and the size of L surrounding T should be large, while N should have similar properties relating to M.

At this level, we thus explored the intra- and inter-slice object-level features. For each merged region R_o , a 32-dimensional feature vector f_o was computed: (Dim. 1-15) the minimum distance from R_o to the type k areas in the d direction (above, below, left, right, and the z direction); (Dim. 16-30) the average size of type k in the d direction relative to R_o , normalized by the dimension of R_o ; (Dim. 31) the size of R_o ; and (Dim. 32) the spatial-level labeling at R_o . Unlike y_l and y_s , the labeling y_o should then take four possible values: L, M, T or N.

Cascaded Learning and Inference. To create the discriminative classifier, we performed piecewise learning for the feature weights α , β and γ (Eq. (1)). We first trained a one-versus-all multi-class support vector machine (SVM) for the local-level model, then another multi-class SVM for the spatial level, and lastly a third one for the object level. At each stage, the training focused on the features of that level only, with classification results of the previous level as the input for feature computation.

Although we could rewrite the score function into structural-SVM type [6], we chose to do SVM-based piecewise learning mainly because: (1) a feature vector combining all three levels generated based on the training data would not capture the cascaded nature of higher-level features dependent on the lower levels, thus would not achieve the optimal performance; and (2) our features were designed to be independent between regions at the same level, so optimization for the entire image collectively was not necessary.

A three-level inference based on mean-shift clustering with the three learned multi-class SVMs was then performed. The final labeling was chosen as the class type with the highest combined margin from three levels. The classification could be done per region using SVM, without considering inter-dependencies between regions, because the spatial relationships were derived based on the labeling of the previous level, not within the same level.

2.2 Pathological Context Description

We described the pathological context for the detected tumor (T) and abnormal lymph nodes (N) in three aspects: (1) texture features: the mean, standard deviation, skewness and kurtosis of the Gabor filtered T and N areas for both CT and PET; (2) shape features: the volume, eccentricity, extent and solidity of T and N; and (3) spatial features: the distance to the chest wall and mediastinum for tumor, and distance to two lung fields for lymph nodes, normalized

by the size of the tumor or lymph node itself. The distances were computed in four directions per slice, and averaged across all slices weighted by the detection score S . So, slices with more obvious T or N regions would contribute more to the spatial feature.

Besides extracting the feature vectors of the detected T/N areas, we also extended the context description with an image retrieval component, to retrieve a set of images with similar pathological patterns for a given query image. The retrieved images, which were stored in the database with diagnosis information, could be used to aid image interpretation. Given the query image I and the image J , the distance was defined as:

$$D_{I,J} = \omega \cdot (|v_I - v_J| / (v_I + v_J)) = \omega \cdot v_{I,J} \quad (2)$$

where v was the feature vector of the image (concatenation of the texture, shape and spatial features of T and N), and ω was the feature weights. A training set was constructed of Q triplets: $\langle I, J, K \rangle$, where I was similar to J , and dissimilar to K . It was thus expected to satisfy $D_{I,K} > D_{I,J}$, and the weight vector ω was computed based on the large-margin optimization method [11]:

$$\operatorname{argmin}_{\omega, \xi \geq 0} \frac{1}{2} \|\omega\|^2 + C \sum_q \xi_q, \quad \text{s.t. } \forall q : \omega \cdot (v_{I,K} - v_{I,J}) \geq 1 - \xi_q \quad (3)$$

The training data $\langle I, J, K \rangle$ captured the search preference, e.g. based on tumor characteristics only, or including lymph nodes. By changing the training data, the derived weights ω would vary and result in different retrievals.

2.3 Materials and Preprocessing

In this study, a total of 1279 transaxial PET-CT image slice pairs were selected from 50 patients with NSCLC. The images were acquired using a Siemens TrueV 64-slice PET-CT scanner (Siemens, Hoffman Estates, IL) at the Royal Prince Alfred Hospital, Sydney. All 50 cases contained primary lung tumors, and 23 of them contained abnormal lymph nodes. The locations of tumors and disease in regional lymph nodes were annotated manually, and for each patient study, the other 49 patient studies were marked similar or dissimilar, as the ground truths. A fully-automatic preprocessing was performed on each CT slice to remove the patient bed and soft tissues outside of the lung and mediastinum, based on simple thresholding, connected component analysis and filling operations. The resulting mask was then mapped to the co-registered PET slice.

3 Results

The structure localization performance for the 50 patient studies is summarized in Table 1a. Based on visual inspections, a volume (case-level) that was classified accurately with its boundary matching closely to the ground truth was considered correct. The multi-level model was trained on slice pairs randomly selected

from 10 imaging studies. To further evaluate the localization performance at the slice level, Table 1b shows the measurements for all 1279 slices. We also compared our method with two other approaches (Table 1c and 1d): a four-class SVM for voxel-level classification; and a four-class SVM for region-level classification after mean-shift clustering (identical to our local-level modeling, except training for four classes). Both approaches were trained using the same set of data as our local-level modeling. Our multi-level modeling showed clear advantages, especially in differentiating tumor and abnormal lymph nodes. As a component of our model, Table 1d illustrated the benefit of region-based processing compared to Table 1c. Some visual results are shown in Figure 1.

Table 1. The pairwise confusion matrix of the four region classes tested on 50 patient studies. (a) Our method - image/case level results. (b) Our method - finer slice-level results. (c) Gabor+SVM - image/case level results. (d) Gabor+Mean-shift+SVM - image/case level results.

Ground Truth	Prediction (%)			
	L	M	T	N
Lung lobe	100	0	0	0
Mediastinum	0	94.3	3.8	1.9
Tumor	0	1.6	84.4	14.1
Lymph node	0	3.7	18.5	77.8

(a)

Ground Truth	Prediction (%)			
	L	M	T	N
Lung lobe	99.2	0.8	0	0
Mediastinum	0	97.1	2.1	0.8
Tumor	1.7	6.1	87.8	4.3
Lymph node	0	7.5	12.3	80.2

(b)

Ground Truth	Prediction (%)			
	L	M	T	N
Lung lobe	100	0	0	0
Mediastinum	13.3	60.2	21.7	4.8
Tumor	6.8	10.2	42.4	40.7
Lymph node	5.6	11.1	27.8	55.6

(c)

Ground Truth	Prediction (%)			
	L	M	T	N
Lung lobe	100	0	0	0
Mediastinum	0	83.3	5.0	11.7
Tumor	0	16.5	43.7	39.8
Lymph node	0	6.5	35.5	58.1

(d)

The sensitivity and specificity of tumor/lymph node localization relative to the lung and mediastinum are listed in Table 2a. In testing, the distances between the tumor and the chest wall and mediastinum/hilum, and between the abnormal lymph nodes and the left and right lung lobes, were assessed to determine the sensitivity and specificity. The remaining errors were mainly caused by misclassifications between tumors near the mediastinum and the abnormal lymph nodes. Our method resulted in higher sensitivity and specificity in deriving the spatial relationships, compared to using only local-level features (Table 2b), because of the highly effective structure localization.

Finally, we evaluated the retrieval performance by using each imaging study as the query to retrieve the most similar cases, and the average precision and recall were computed. We compared our method with techniques based on weighted histogram and bag-of-SIFT [12] features for global and local feature extraction; and both approaches were trained in the same way as our method for similarity measure. As shown in Table 3, our method achieved much higher precision and

Table 2. The sensitivity (SE) and specificity (SP) of the tumor and lymph node localization relative to the lung lobes and mediastinum tested on 50 cases. (a) Our method and, (b) Gabor+Mean-shift+SVM.

	Tumor		Lymph node	
	Wall	Hilum	Left	Right
SE (%)	100	97.2	92.9	88.9
SP (%)	98.0	84.8	89.4	91.5

(a)

	Tumor		Lymph node	
	Wall	Hilum	Left	Right
SE (%)	83.3	82.9	93.3	87.5
SP (%)	98.0	77.8	65.0	69.2

(b)

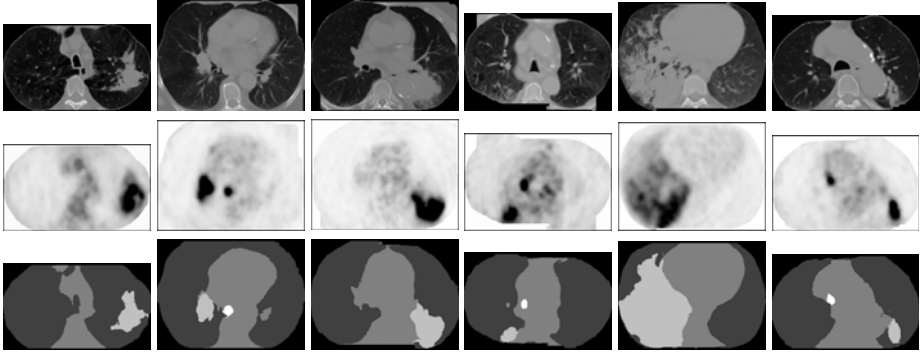


Fig. 1. Six examples of structure localization, showing one transaxial slice pair per case. The top row is the CT image slice (after preprocessing); the middle row is the co-registered PET slice; and the bottom row shows the localization results, with 5 different gray scale values (black to white) indicating background, L, M, T and N.

Table 3. The precision-recall measure of the retrieval results of the top one, three or five most similar matches on 50 cases. Our method is compared with the histogram (HIST) and bag-of-SIFT [12] features (BoSF) based approaches.

	Precision (%)			Recall (%)		
	Ours	HIST	BoSF	Ours	HIST	BoSF
Top-1	84.0	46.0	32.0	14.1	7.5	6.0
Top-3	70.7	34.7	29.3	31.4	14.1	12.7
Top-5	63.2	28.8	25.6	44.3	19.7	17.9

recall. The results showed that our method could extract the salient (pathological) features more effectively than the general techniques; and suggest that the detected context could be used in a CBIR system.

4 Conclusions

We proposed a new method to automatically detect the primary tumor and disease in lymph nodes, and the spatial relationships with the lung and

mediastinum on PET-CT thoracic images. By exploring a comprehensive set of features at the local, spatial and object levels, the discriminative classification achieves an accurate localization of the various structures in the thorax. The work is an initial step towards a computer aided system for PET-CT imaging diagnosis for lung cancer staging. The extracted pathological contexts also show high precision when used to retrieve the most similar images.

References

1. Tao, Y., Lu, L., Dewan, M., Chen, A.Y., Corso, J., Xuan, J., Salganicoff, M., Krishnan, A.: Multi-level ground glass nodule detection and segmentation in CT lung images. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009, Part II. LNCS, vol. 5762, pp. 715–723. Springer, Heidelberg (2009)
2. Kakar, M., Olsen, D.R.: Automatic segmentation and recognition of lungs and lesions from CT scans of thorax. *Comput. Med. Imag. Graph.* 33(1), 72–82 (2009)
3. Feulner, J., Zhou, S.K., Huber, M., Hornegger, J., Comaniciu, D., Cavallaro, A.: Lymph nodes detection in 3-D chest CT using a spatial prior probability. In: CVPR, pp. 2926–2932 (2010)
4. Wojak, J., Angelini, E.D., Bloch, I.: Joint variational segmentation of CT-PET data for tumoral lesions. In: ISBI, pp. 217–220 (2010)
5. Wu, D., Lu, L., Bi, J., Shinagawa, Y., Boyer, K., Krishnan, A., Salganicoff, M.: Stratified learning of local anatomical context for lung nodules in CT images. In: CVPR, pp. 2791–2798 (2010)
6. Desai, C., Ramanan, D., Fowlkes, C.: Discriminative models for multi-class object layout. In: ICCV, pp. 229–236 (2009)
7. Galleguillos, C., McFee, B., Belongie, S., Lanckriet, G.: Multi-class object localization by combining local contextual interactions. In: CVPR, pp. 113–120 (2010)
8. Shotton, J., Winn, J., Rother, C., Criminisi, A.: TextonBoost: joint appearance, shape and context modeling for multi-class object recognition and segmentation. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006, Part I. LNCS, vol. 3951, pp. 1–15. Springer, Heidelberg (2006)
9. Comaniciu, D., Meer, P.: Mean shift: a robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* 24(5), 603–619 (2002)
10. Song, Y., Cai, W., Eberl, S., Fulham, M.J., Feng, D.: Automatic detection of lung tumor and abnormal regional lymph nodes in PET-CT images. *J. Nucl. Med.* 52(supplement. 1), 211 (2011)
11. Frome, A., Singer, Y., Sha, F., Malik, J.: Learning globally-consistent local distance functions for shape-based image retrieval and classification. In: ICCV, pp. 1–8 (2007)
12. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *Int. J. of Comput. Vis.* 60(2), 91–110 (2004)

X-ray Categorization and Spatial Localization of Chest Pathologies

Uri Avni¹, Hayit Greenspan^{1,*} and Jacob Goldberger²

¹ BioMedical Engineering Tel-Aviv University, Israel

² School of Engineering, Bar-Ilan University, Israel

Abstract. In this study we present an efficient image categorization system for medical image databases utilizing a local patch representation based on both content and location. The system discriminates between healthy and pathological cases and indicates the subregion in the image that is automatically found to be most relevant for the decision. We show an application to pathology-level categorization of chest x-ray data, the most popular examination in radiology. Experimental results are provided on chest radiographs taken from routine hospital examinations.

Keywords: Image categorization, x-ray, chest radiography, visual words, Computer-Aided Diagnosis (CAD), region-of-interest (ROI).

1 Introduction

In the last ten years the number of images that are acquired every day in any modern hospital has increased exponentially, due to the progress in digital medical imaging techniques and patient image-screening protocols. One immediate consequence of this trend is the enormous increase in the number of images that must be reviewed by radiologists. This has led to a concomitant demand for computerized tools to aid radiologists in the diagnostic process. Computerized systems assist the radiologist in the diagnostic process by categorizing the image content. This is done by learning from a large archive of image examples that are annotated by experts.

Image categorization is concerned with the labeling of images into predefined classes. The principal challenge of image categorization is the capture of the most significant features within the images that facilitate the desired classification. A single image can contain a large number of regions-of-interest (ROI), each of which may be the focus of attention for the medical expert, depending on the diagnostic task at hand. A single chest image for example, contains the lungs, heart, rib cage, diaphragm, clavicle, shoulder blade, spine and blood vessels, any of which may be the focus of attention. One way to enhance the image categorization process is to focus on a ROI within the image that is relevant to the presumed pathology. The advantage of the ROI approach is that the rest

* Currently on Sabbatical at IBM Almaden Research Center, San Jose, CA.

of the image can be ignored leading to computational advantages and increased accuracy in the classification.

Clinical decision support techniques based on automatic classification algorithms can produce a strong need to localize the area that is most relevant to the diagnostic task. A diagnostic system that, in addition to a decision on the existence of a pathology, can provide the image region that was used to make the decision can assist radiologists to better understand the decision and evaluate its reliability. Another advantage of an ROI based decision is that we can construct a detailed representation of the local region. We refer to this approach as ROI based image categorization. Of course it is not suited to every image categorization task; not every pathology includes a significant and identifiable ROI that appears across the data set. In such cases, a global, full-image categorization is appropriate.

The (bags of) visual words paradigm, which has recently been adapted from the text retrieval domain to the visual analysis domain (see e.g. [1][2]), provides an efficient way to address the medical imaging categorization challenge in large-size archives while maintaining solid classification rates [3][4]. The best categorization methods in recent ImageCLEF competitions are all based on variants of the visual words concept [5]. In this study we utilize a variant of the visual words framework, that combines content and location, to automatically localize the relevant region for a given diagnostic task. Besides localizing the decision based area, the proposed method yields improved results over a categorization system based on the entire image.

2 The Visual Words Framework for Classification

In this section we describe a state-of-the-art medical image categorization paradigm using the visual words framework, which is based on a large set of image patches, and their respective representation via a learned dictionary. This paradigm is the foundation for the proposed localized image classification system. The method was ranked first in the ImageCLEF 2009 medical annotation task [6].

Given a training labeled image dataset, patches are extracted from every pixel in the image. Each small patch shows a localized view of the image content. In the visual dictionary learning step, a large set of images is used (ignoring their labels). We extract patches using a regular grid, and normalize each patch by subtracting its mean gray level, and dividing it by its standard deviation. This step insures invariance to local changes in brightness, provides local contrast enhancement and augments the information within a patch. Patches that have a single intensity value are ignored in x-ray images (e.g. the brightness of the air surrounding the organ appears uniform especially in DICOM format). We are left with a large collection of several million vectors. To reduce both the computational complexity of the algorithm and the level of noise, we apply a Principal Component Analysis procedure (PCA) to this initial patch collection. The first few components of the PCA, which are the components with the largest eigenvalues, serve as a basis for the information description. In addition to the patch content information represented, we add the patch center coordinates to the feature vector. This introduces spatial information into the image representation,

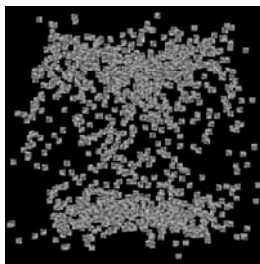


Fig. 1. An example of a region based visual dictionary. Each visual word is placed according to its (x, y) coordinates.

without the need to explicitly model the spatial dependency between patches. The final step of the visual-words model is to convert vector-represented patches into *visual words* and generate a representative *dictionary*. A visual word can be considered to be a representative of several similar patches. The vectors are clustered into k groups in the feature space using the k -means algorithm. The resultant cluster centers serve as a vocabulary of k visual words. The location of the cluster center is the average location of all the patches in the cluster. Due to the fact that we included spatial coordinates as part of the feature space, the visual words have a localization component in them, which is reflected as a spatial spread of the words in the image plane. Words are denser in areas with greater variability across images in the database. Fig. 1 shows a region based visual dictionary. Each visual word is placed according to its (x, y) coordinates.

A given (training or testing) image can now be represented by a unique distribution over the generated dictionary of words. In our implementation, patches are extracted from every pixel in the image. For a 512×512 image, there are several hundred thousand patches. The patches are projected into the selected feature space, and translated (quantized) to indices by looking up the most similar word in the generated dictionary. Note that as a result of including spatial features, both the local content and spatial layout of the image are preserved in the discrete histogram representation. Image classification of a test image is based on the ground truth of manually categorized images that are used to train an SVM classifier which is applied to the image histogram representation.

3 Localizing Image Categorization

When radiologists look for a specific pathology in an image, they usually focus on a certain area. For example right pleural effusion is diagnosed using the lower bottom and the peripheral lateral zones of the right lung, while ignoring the rest of the chest x-ray image. In this section we describe how to automatically find a relevant area, without prior knowledge about the organ structure or the characteristics of the pathology. This step is designed to improve the classification accuracy of the global approach. It can also provide a useful visualization of the area used in the automatic classification.

Assume we already learned a visual dictionary as described in the previous section and we are now concentrating on a specified pathology. We are given a training image set in which each image is manually labeled as either healthy or pathological. The visual-words representation of an image x is a histogram vector (x_1, \dots, x_k) such that x_i is the relative number of image patches in x that were mapped to the i -th visual word based on content and location similarity. These k numbers are the features extracted from the image to be used in the classification algorithm, where each feature corresponds to a visual word. The first step toward localization of the pathology decision is finding the relevance of each feature. Feature relevance is often characterized in terms of mutual information between the feature and the class label. It is reasonable to expect that for a feature (visual word) that is located far from the pathology area, the class label and the feature values random variables should be independent. We compute the mutual information between the image label and each of the features in the following way. Suppose we are given n images with binary (healthy/pathological) labels c_1, \dots, c_n and the feature vector of the t -th image is denoted by (x_{t1}, \dots, x_{tk}) . To obtain a reliable estimation of the mutual information for each feature i , we first quantize the i -th feature values across all the images x_{1i}, \dots, x_{ni} into L levels (in our implementation we sort the n values and divide them into four groups of equal size). Denote the quantized version of x_{ti} by $y_{ti} \in \{1, \dots, L\}$. Denote the joint probability of the (quantized) i -th feature and the image class by:

$$p_i(v, c) = \frac{1}{n} |\{t | y_{ti} = v, c_t = c\}| \quad (1)$$

where c is the class label, v is the quantized bin level and $|\cdot|$ is the set cardinality. The mutual information between the class label variable C and the quantized feature Y_i is [7]:

$$I(Y_i; C) = \sum_{v=1}^L \sum_c p_i(v, c) \log \frac{p_i(v, c)}{p_i(v)p(c)} \quad (2)$$

where $p_i(v)$ and $p(c)$ are the marginal distributions of the i -th feature and the class label respectively. C is a binary random variable and therefore $0 \leq I(Y_i; C) \leq 1$.

Up to now we have computed the relevance of each feature (visual word) separately. However, since each visual word has a location, we can consider the relevance of an entire region as the relevance of all the features located in that region. The proposed method can be viewed as a filter-based feature selection. Unlike general feature selection problems, here the features are located in the image plane. Hence, instead of feature selection we apply region selection. Since the visual words have a spatial location, the relevance information can be represented in the image plane. We create a relevance map $R(x, y) = \sum_i I(Y_i, C)$ such that i goes over all the visual words that are located at (x, y) . $R(x, y)$ is a matrix with mutual information values at the positions of the visual-word feature centers, and zero at other locations. In this representation, areas that contain highly relevant features are likely to be the regions of interest for the classification task.

We define the relevance of the rectangular $[x_1, x_2] \times [y_1, y_2]$ sub-region in the image plane to the pathology as

$$\text{score}(x_1, x_2, y_1, y_2) = \sum_{x_1 \leq x \leq x_2} \sum_{y_1 \leq y \leq y_2} R(x, y) = \sum_i I(Y_i, C) \quad (3)$$

such that i goes over all the visual words that are located in the rectangle $[x_1, x_2] \times [y_1, y_2]$. We look for a rectangular region that contains the maximum amount of relevant features. For a given region size, we examine all the rectangles in the image and look for the rectangle with the highest score. This search can be efficiently done using the integral image algorithm which is part of the seminal Viola-Jones object detection framework [8]. The integral image algorithm efficiently generates the sum of values in rectangular subsets of a grid. It can be computed in one pass over the image using the following recurrence relation:

$$\begin{aligned} \text{sI}(x, y) &= \text{sI}(x, y - 1) + R(x, y) \\ \text{II}(x, y) &= \text{II}(x - 1, y) + \text{sI}(x, y) \end{aligned} \quad (4)$$

The region score is thus:

$$\text{score}(x_1, x_2, y_1, y_2) = \text{II}(x_2, y_2) - \text{II}(x_1, y_2) - \text{II}(x_2, y_1) + \text{II}(x_1, y_1) \quad (5)$$

The relevance map $R(x, y)$ takes into account all of the training images and therefore the ROI is little affected by noisy images. However, since the ROI is calculated globally, this procedure finds a rough estimation of the region of interest. The exact region might vary between images. The ROI can be refined individually for each image by examining the mutual information map in the image space instead of the visual dictionary space. Every patch in the image is translated into a visual word. We can create a relevance map per image $R_{image}(x, y)$ by placing in each image patch center the mutual information of the visual word it is assigned to. In other words, $R_{image}(x, y) = I(Y_i, C)$ such that i is the visual word that the image patch centered at (x, y) is assigned to. We can then repeat the integral-image process on the map R_{image} to find a smaller rectangle with the maximal amount of relevant patches in the image relevance map. The search is confined to the rough ROI found in the first step. A two-step process is required because if the relevance map of the images is noisy, it may generate erroneous ROI, especially if its relevant area is relatively small.

After finding an ROI and refining it for each image in the healthy/pathological labeled training set, we run a second training stage, where sub images are cropped to the region of interest of the pathology. A new dictionary is generated for each pathology, and the SVM classifiers are trained using the words histograms from the cropped regions.

In the test phase, a new image is converted to a word histogram using the dictionary learned in the first step, and an image relevance map is calculated. For each pathology we crop the image using the rough global ROI that was found in the training phase. Next we find a refined ROI by applying the integral image algorithm to the test image relevance map R_{image} . The refined ROI is then

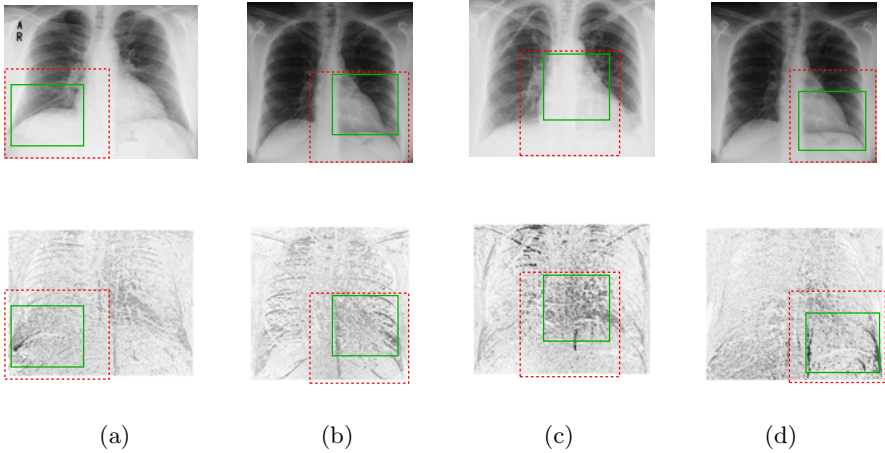


Fig. 2. Frontal chest x-ray images, top row shows the ROI detection overlaid on the original image and bottom row shows the corresponding relevance map. Global ROI is shown in red and the refined ROI in green. The pathologies are (a) right pleural effusion, (b) enlarged heart shadow, (c) enlarged mediastinum and (d) left pleural effusion.

converted to a words histogram using the dictionary from the second training step, and passed to the region based healthy/pathology classifier that was also trained in the second training step. The reported result is the binary decision and the image subregion that was used to obtain the decision. Fig. 2 shows examples of relevance maps and ROI detection (both global for a pathology and image refined) of chest x-ray images. The entire processing of a test image including the translation to visual words and the integral image computation takes less than a second (time was measured on a dual quad-core Intel Xeon 2.33 GHz.).

4 Experiments

Chest radiographs are the most common examination in radiology. They are essential for the management of various diseases associated with high mortality and morbidity and display a wide range of potential information, many of which are subtle. According to a recent survey [9], most of research in computer-aided detection and diagnosis in chest radiography has focused on lung nodule detection. However, lung nodules are a relatively rare finding in the lungs. The most common findings in chest x-rays include lung infiltrates, catheters and abnormalities of the size or contour of the heart [9]. Distinguishing the various chest pathologies is a difficult task even for human observers. Research is still needed to develop an appropriate set of computational tools to support this task. We used 443 frontal chest images in DICOM format from the Sheba medical center hospital PACS, taken during routine examinations. X-ray interpretations,

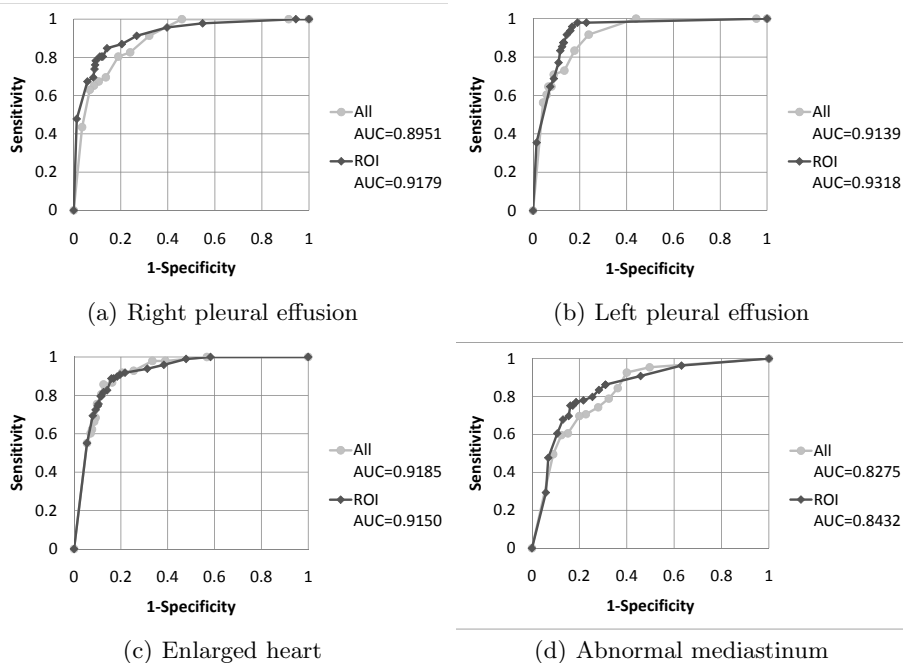


Fig. 3. ROC curves for pathology detection in frontal chest x-rays, using the entire image and automatically detected ROI; the areas under the curves (AUC) were calculated using trapezoidal approximation

made by two radiologists, served as the reference gold standard. The radiologists examined all of the images independently; they then discussed and reached a consensus regarding the label of every image. The pathology data include 98 images with enlarged heart shadow, 109 images with enlarged mediastinum, 46 images with right pleural effusion and 48 images with left pleural effusion. Some patients had multiple pathologies.

The original high-resolution DICOM images were initially resized to a maximal image dimension of 1024 pixels, with aspect-ratio maintained. We followed the method described in Section 2 to extract features, build a visual dictionary, and represent an image as a histogram of visual words. To avoid overfitting and to preserve the generalization ability of the classifiers, model parameters were chosen following the experiments on the ImageCLEF database, described in [6, 10]. For each pathology we found the relevant ROI and utilized the cropped images (both healthy and pathological) to learn a visual dictionary. The rough ROI size was selected to be 36% of the image area; it was cropped to a smaller rectangle if it passed the image border. The fine ROI was set to 15% of the image area. We then detected each of the four pathologies using a binary SVM classifier with a χ^2 kernel, trained on words histogram built from the fine ROI. The sensitivity and specificity were calculated using leave-one-out cross validation.

Modifying the relative cost of false negative errors in the SVM cost minimization problem determines the tradeoff point between sensitivity and specificity. This technique was used to produce the receiver operating characteristic (ROC) curves for several pathologies, shown in Fig. 3. The figure clearly indicates that in three out of the four pathologies our localized categorization method outperformed the global categorization that utilizes patches from the entire image. In the right pleural effusion case, the AUC is improved from 0.895 to 0.92; In the left pleural effusion case, the AUC is improved from 0.91 to 0.93, and in the abnormal mediastinum case, an improvement in the AUC is from 0.827 to 0.84.

To conclude, in this study we showed how visual word information, based on both content and location, can be used to automatically localize the decision region for pathology detection. The method is based on choosing the visual words that are most correlated with the diagnoses task. We have shown that the proposed method outperforms methods that are based on the entire image, both in terms of classification performance and in enabling human interpretation of the decision. The method proposed is general and can be applied to additional Chest x-ray pathologies, currently being explored, as well as to additional medical domains.

Acknowledgments. The authors wish to thank Dr. Eli Konen and Dr. Michal Sharon from the Diagnostic Imaging Department of Sheba Medical Center, Tel Hashomer, Israel - for the data collection, guidance and helpful discussions.

References

1. Varma, M., Zisserman, A.: Texture classification: are filter banks necessary? In: CVPR, vol. 2, pp. 691–698 (2003)
2. Fei-Fei, L., Perona, P.: A bayesian hierarchical model for learning natural scene categories. In: CVPR (2005)
3. Deselaers, T., Hegerath, A., Keysers, D., Ney, H.: Sparse patch-histograms for object classification in cluttered images. In: Franke, K., Müller, K.-R., Nickolay, B., Schäfer, R. (eds.) DAGM 2006. LNCS, vol. 4174, pp. 202–211. Springer, Heidelberg (2006)
4. Tommasi, T., Orabona, F., Caputo, B.: Discriminative cue integration for medical image annotation. *Pattern Recogn. Lett.* 29(15), 1996–2002 (2008)
5. Muller, H., Clough, P., Deselaers, T.: *Imageclef: Experimental evaluation in visual information retrieval*. Springer, Heidelberg (2010)
6. Avni, U., Goldberger, J., Greenspan, H.: Addressing the imageclef, challenge using a patch-based visual words. In: *The Cross Language Evaluation Forum Workshop (CLEF)* (2009)
7. Cover, T.M., Thomas, J.A.: *Elements of Information Theory*. Wiley-Interscience, Hoboken (1991)
8. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: *IEEE Computer Vision and Pattern Recognition* (2001)
9. van Ginneken, B., Hogeweg, L., Prokop, M.: Computer-aided diagnosis in chest radiography: Beyond nodules. *European Journal of Radiology* 72(2), 226–230 (2009)
10. Avni, U., Greenspan, H., Konen, E., Sharon, M., Goldberger, J.: X-ray categorization and retrieval on the organ and pathology level, using patch-based visual words. *IEEE Trans. Medical Imaging*, 733–746 (2011)

Computer-Aided Detection of Ground Glass Nodules in Thoracic CT Images Using Shape, Intensity and Context Features

Colin Jacobs^{1,2}, Clara I. Sánchez², Stefan C. Saur³, Thorsten Twellmann³, Pim A. de Jong⁴, and Bram van Ginneken^{2,5}

¹ Fraunhofer MEVIS, Bremen, Germany

² Diagnostic Image Analysis Group, Radboud University Nijmegen Medical Centre, The Netherlands

³ MeVis Medical Solutions AG, Bremen, Germany

⁴ Department of Radiology, University Medical Center, Utrecht, The Netherlands

⁵ Image Sciences Institute, University Medical Center, Utrecht, The Netherlands

Abstract. Ground glass nodules (GGNs) occur less frequent in computed tomography (CT) scans than solid nodules but have a much higher chance of being malignant. Accurate detection of these nodules is therefore highly important. A complete system for computer-aided detection of GGNs is presented consisting of initial segmentation steps, candidate detection, feature extraction and a two-stage classification process. A rich set of intensity, shape and context features is constructed to describe the appearance of GGN candidates. We apply a two-stage classification approach using a linear discriminant classifier and a GentleBoost classifier to efficiently classify candidate regions. The system is trained and independently tested on 140 scans that contained one or more GGNs from around 10,000 scans obtained in a lung cancer screening trial. The system shows a high sensitivity of 73% at only one false positive per scan.

Keywords: ground glass nodule, computer-aided detection, chest CT.

1 Introduction

Ground glass nodules (GGNs) are relatively rare findings in chest computed tomography (CT) examinations. These nodules have an increased attenuation but do not, like solid nodules, completely obscure the lung parenchyma [1], although they may have a solid component. It has been shown that GGNs have a much higher chance of being malignant than solid nodules [1]. Early detection of GGNs is therefore highly important. Beigelman-Aubry et al. [2] showed that both radiologists and computer-aided detection (CAD) systems designed for solid nodules have difficulties with detecting GGNs.

At present, no complete CAD system for GGNs has been tested on a large database. Kim et al. [3] described a slice-based CAD system using texture and intensity features that had a high false positive rate (FPR) and that was tested on only 14 patients. Zhou et al. [4] developed an automatic scheme for both

detection and segmentation of GGNs based on vessel suppression, intensity and texture analysis. They reported high performance but the test data set contained only 10 GGNs. Ye et al. [5] presented a voxel-based method with rule-based filtering that was tested on 50 CT examinations with 52 GGNs. They reported a high sensitivity of 92.3% but also a high FPR of 12.7 per scan. Tao et al. [6] developed a multi-level detection scheme with classification at voxel-level and object-level. They focused on classification of small volumes of interest (VOIs) generated by a candidate detector algorithm which was not otherwise specified. The method was tested on a set of 1100 VOIs including 100 positive ones, from 153 healthy and 51 diseased patients. Results were provided for VOIs only, and neither the FPR per scan nor the total number of VOIs per scan were reported.

In this work, we focus on the automated detection of GGNs from thin-slice CT examinations. In contrast to the aforementioned works, the CAD system is tested on a large data set. A complete detection pipeline is presented, consisting of initial segmentation steps, candidate detection, feature extraction and classification. A comprehensive set of intensity and shape features are computed for each candidate region. As previous studies reported [3,7,8], false positive findings arise from partial volume effects of bronchovascular bundles, the chest wall, the dome of the diaphragm and large vessels. Therefore, we include context features that describe the position of the candidate region with respect to surrounding objects such as airways and the lung boundary.

2 Data and Experimental Design

Data for this study was provided by a large multi-center lung cancer screening trial with thin-slice, low-dose CT scans of current and former heavy smokers. We collected all CT examinations between April 2004 and April 2009 from one screening site, totaling around 10,000 scans from around 3,000 participants. From this data set, all scans in which at least one GGN was reported were selected. This resulted in 140 scans from 58 patients, including 76 unique GGNs. We considered GGNs in follow-up examinations as separate GGNs, leading to a total of 176 GGNs. For each GGN, a manual segmentation was provided. The effective diameter of the GGNs varied from 3.9 to 29.7 mm (median 13.9 mm). All CT examinations were performed with a slice thickness of 0.7 mm and the in-plane voxel size varied between 0.52 and 0.84 mm.

The data set was randomly split into two sets on a patient level, preventing data from the same patient being present in both sets. The training set consisted of 67 scans with 91 GGNs from 31 patients. The test set of 73 CT examinations with 85 GGNs from 27 patients was not touched during system development and was only used for evaluation of the final configuration of the CAD system.

3 Methods

Prior to candidate extraction, we apply a previously developed lung and airway segmentation algorithm [9,10] to each scan.

3.1 Initial Candidate Detection

The candidate detection procedure starts with applying a double-threshold density mask within the lung regions to obtain voxels with attenuation values defined as ground glass opacity. In this study, we used a range between -750 and -300 Hounsfield units (HU) [7,8]. At the boundaries of the lungs, vessels, and airways, partial volume effects lead to attenuation values in the defined range. Therefore, we apply a morphological opening operation using a spherical structuring element with a diameter of 3 voxels to remove the voxels at these edges from the density mask. After this, connected component analysis is performed to obtain candidate regions. Evidently, this results in a large amount of candidate regions. We eliminate all candidate regions which have a volume smaller than 34 mm^3 (volume of an ideal sphere with diameter of 4 mm). Current clinical guidelines [11] state that GGNs smaller than 5 mm do not require follow-up CT examinations and since volume measurements on CT have a certain variability due to partial volume effects, a safety margin of 1 mm is used in this system.

3.2 Features

We defined a rich set of features that can be subdivided into three categories:

Intensity features. The first group of intensity features consists of histogram statistics computed on a normalized histogram with a bin size of 1 HU. Four histograms are constructed from: voxels within the candidate mask, voxels within the bounding box defined by the candidate mask, voxels in the neighborhood created by dilating the candidate mask with a rectangular structuring element of size $3 \times 3 \times 3$ voxels and similarly, but using a rectangular structuring element of size $5 \times 5 \times 5$ voxels. The following histogram statistics are extracted: entropy, mean, height of mean bin, mode, height of mode bin and quantiles at 5%, 25%, 50%, 75% and 95%. Furthermore, we calculate the mean, standard deviation, minimum, maximum and the first 7 invariant Hu moments [12] over the intensity values of voxels within the candidate mask. Local binary patterns (LBP) [13] and Haar wavelets are used for texture analysis. Local binary patterns are computed from the bounding box defined by the candidate mask in which we resample this area to respectively a $16 \times 16 \times 16$ and $32 \times 32 \times 32$ volume. We apply the same histogram statistics to the histogram output of the LBP operator and use these as features. Using 2D Haar wavelets, all axial slices of the $32 \times 32 \times 32$ resampled volume are decomposed into four bands. Then, all bands of the 32 slices are combined and histogram statistics are extracted from the horizontal, vertical and diagonal component of the combined high-frequency part. Finally, maximum vesselness [14] over multiple scales (1.0, 1.77, 3.16, 5.62, and 10.0 voxels) is computed for the voxels in the candidate mask and the mean and standard deviation of the vesselness values are used as features. The total number of intensity features is 103.

Shape features. Shape analysis of the candidate regions is performed using the binary mask of the candidate region. We calculate sphericity, compactness and

volume of the candidate region. In order to calculate the sphericity, we define a sphere S at the center of mass of the candidate region which has equal volume as the candidate region. Then, sphericity is defined as the ratio between the volume of the voxels of the candidate region within sphere S and the total volume of sphere S . For compactness, we used the ratio between the surface of the candidate region and its volume. Furthermore, the same set of 7 invariant Hu moments are computed from the candidate mask voxels to describe its appearance. Note that in contrast to the previous calculation of Hu moments, the voxels are in this case not weighed by their intensity value. This results in 10 shape features.

Context features. In the third category of features, the location of the candidate region in respect to the lung boundary and the airway tree is computed. For all voxels inside the candidate segmentation, the distance to the lung boundary and distance to the closest airway is computed. Then, the mean, standard deviation, minimum and maximum distance to the lung boundary and airways are computed and used as context features. Finally, using the lung segmentation, a bounding box is defined around the lungs. Using this bounding box, relative position features are computed, including relative X, Y and Z position, distance to center of mass of both lungs and distance to left bottom corner of bounding box. This yields 13 context features.

3.3 Classification

In the classification step, candidate regions are classified into GGN or non-GGN class using a two-stage classification approach. Note that in the second stage, the posterior probability of the first classifier is added as an additional feature.

In pilot experiments, we extensively tested different classifiers (Linear Discriminant Analysis (LDA), k-Nearest Neighbor (kNN) and GentleBoost [15]) for the first and second phase classification. In these experiments, 10-fold cross-validation on the training set was performed to test the performance of the different classifiers. Note that the 10 folds were again created by splitting the training set at a patient level. Consequently, all follow-up examinations of one patient were in the same fold to prevent bias.

For the first phase, we ranked all features according to Fisher’s discriminant ratio [16] and we selected the four features with the highest ranking. Using these four features (two shape and two intensity features), LDA and kNN were tested and LDA proved to give slightly better results. Using the results from the 10-fold cross-validation on the training set, the posterior probability threshold for the first phase classification was determined. The threshold was set such that no true positives were lost in the first phase for the training set. This reduced the number of candidates in the training set by 66%. Consequently, all features only need to be calculated for about one third of all candidate regions, which accelerates the CAD system considerably ($\sim 40\%$).

For the second phase classification, we experimented with an LDA, kNN, and GentleBoost classifier. Optimal results for the kNN-classifier were found using $k=60$ and we used regression stumps as a weak classifier for the GentleBoost classification. Since the data set consists of a relatively high amount of features

Table 1. Performance of the CAD system for different feature groups. Sensitivity is reported at $\frac{1}{8}$, $\frac{1}{4}$, $\frac{1}{2}$, 1, 2, 4, and 8 false positives per scan. The score is the average sensitivity at these 7 operating points.

Feature set	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{2}$	1	2	4	8	Score
All features	0.437	0.565	0.671	0.729	0.748	0.780	0.788	0.674
Shape and intensity features	0.476	0.569	0.650	0.686	0.705	0.765	0.772	0.660
Context and intensity features	0.465	0.578	0.645	0.687	0.710	0.749	0.767	0.657
Context and shape features	0.479	0.521	0.595	0.683	0.737	0.778	0.780	0.654
Intensity features	0.497	0.575	0.630	0.668	0.700	0.723	0.731	0.647
Context features	0.402	0.472	0.569	0.627	0.706	0.720	0.724	0.603
Shape features	0.486	0.506	0.550	0.604	0.637	0.708	0.724	0.602

(127) and a smaller amount of true positives (around 100), we decided to select the best 20 features using a Sequential Feed Forward Selection (SFFS) procedure to prevent overfitting of the classifier. During the SFFS procedure, the partial area under the curve (AUC) of the ROC curve was used as objective function. The upper threshold on false positive fraction was set at the value which corresponds to 5 false positives per scan. As the concept of boosting is based on sequentially applying weak classifiers on a subset of the data [15], feature selection was not used for the GentleBoost classifier. Finally, after extensive testing of different classifiers with combined feature selection, we concluded that the GentleBoost classifier had a slightly better performance and therefore we used this in the final configuration of the system.

In some cases multiple candidate regions were present for a single GGN. As we focused on detection, we counted a GGN as detected when at least one matching candidate was classified as positive. The remaining matching candidates were considered neutral in the evaluation and not counted as false negatives.

4 Results

The candidate detection step generated 524 ± 308 candidate regions per scan. Candidates are considered positive when the centers of mass of the GGN segmentation and the candidate region were within a distance d of each other. For $d = 10$ mm, the sensitivity of the candidate detector was 92% (84/91) and 87% (74/85) for the training and test set, respectively. For sake of readability, we omit a detailed reporting for a distance criterion of $d = 5$ mm where the results were comparable to the ones of $d = 10$ mm.

The first stage LDA classifier and second phase GentleBoost classifier were trained with all candidates from the training set and tested on the test set. After the first classification step, 32% of the candidate regions remained in the test set at the expense of eliminating three true positives. The FROC curve of the complete CAD system is given in Fig. 1 and sensitivities at various operating points are given in Table 1. At only one false positive per scan, 62 out of 85 GGNs (73% sensitivity) were detected.

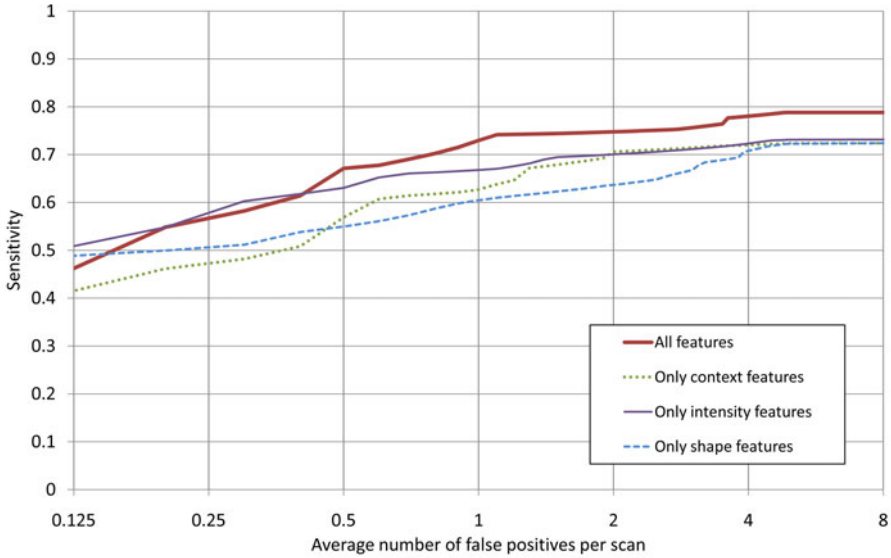


Fig. 1. FROC curves with a logarithmic x-axis. Results are shown for the proposed CAD system and systems that are only trained with one type of features.

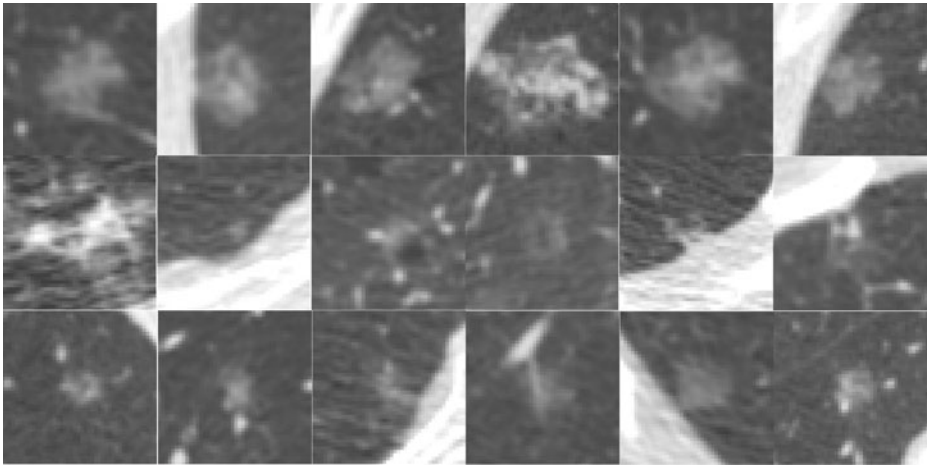


Fig. 2. Examples of true positives, false negatives and false positives of the CAD system. All images are axial views of 30×30 mm with a window level of $-600/1600$ HU. The top row shows the six true positives with the highest degree of suspicion in the test set according to the CAD system. The middle row shows six false negatives that the CAD system did not detect when set to operate at 1 false positive per scan. These nodules were picked up by the candidate detector, but they were not deemed suspicious enough. Finally, the bottom row shows the six false positives with the highest degree of suspicion in the test set according to the CAD system.

Furthermore, we investigated the effect of the three separate feature groups by using only one or only two in the second phase classification. Results are given in Fig. 1 and Table 1. An overall performance metric is derived from the FROC curves by averaging the sensitivities at $\frac{1}{8}$, $\frac{1}{4}$, $\frac{1}{2}$, 1, 2, 4, and 8 false positives per scan. These results indicate that the combination of all features increases the performance considerably. In Fig. 2, we show examples of ground glass nodules which were correctly detected and missed by the CAD system.

5 Discussion and Conclusion

The FROC curve in Fig. 1 shows that our CAD system is able to find around 50% of all GGNs without any false positives. This is a very encouraging result. Moreover, a radiologist who retrospectively inspected the false positives of the system operating at 1 FP/scan (see Fig. 2, bottom row) indicated that many of these seemed to be GGNs. The reason why these findings had not been annotated may be that the scans in which these findings occurred contained multiple areas of ground glass opacity with a pattern resembling smoking related interstitial lung disease. On the other hand it may be that these were real GGNs missed by human readers. This interesting observation does require further study.

In clinical practice, a GGN CAD system will be used in combination with a solid nodule CAD system, which also produces false positives, and therefore we believe that the operating point at only 1 FP/scan is optimal. Moreover, the sensitivity does not increase much after 1 FP/scan.

From the 22 missed GGNs, 11 were actually missed in the candidate extraction step. We observed that in some cases a single GGN was detected as two separate candidate regions that were subsequently eliminated. Possible improvements are candidate clustering or integrating voxel classification in the candidate detection procedure, as done in [6]. Furthermore, even though the data for this study originated from over 10,000 scans obtained in a screening trial, the training set still contained less than 100 GGNs. We plan to collect a larger and more diverse training database in the future. This may help the CAD system to also recognize uncommon manifestations of GGNs, such as the ones on the second row of Fig. 2.

In conclusion, a complete computer-aided detection scheme for detection of ground glass nodules has been presented and tested on a large database. A comprehensive set of intensity, shape and context features was used to describe the appearance of a ground glass nodule. An optimized classification scheme using two stages of classification was employed. We evaluated the performance of the CAD system on an independent test set that was not touched during system development and obtained a sensitivity of 73% at only one false positive detection per scan. This is a substantially better performance than reported in previous work [3,6,4,5]. We are convinced that this performance level is sufficient for application of the system in clinical practice.

References

1. Henschke, C.I., Yankelevitz, D.F., Mirtcheva, R., McGuinness, G., McCauley, D., Miettinen, O.S.: CT screening for lung cancer: Frequency and significance of part-solid and nonsolid nodules. *AJR Am. J. Roentgenol.* 178(5), 1053–1057 (2002)
2. Beigelman-Aubry, C., Hill, C., Boulanger, X., Brun, A., Leclercq, D., Golmard, J., Grenier, P., Lucidarme, O.: Evaluation of a computer aided detection system for lung nodules with ground glass opacity component on multidetector-row CT. *J. Radiol.* 90(12), 1843–1849 (2009)
3. Kim, K.G., Goo, J.M., Kim, J.H., Lee, H.J., Min, B.G., Bae, K.T., Im, J.G.: Computer-aided diagnosis of localized ground-glass opacity in the lung at CT: Initial experience. *Radiology* 237, 657–661 (2005)
4. Zhou, J., Chang, S., Metaxas, D.N., Zhao, B., Ginsberg, M.S., Schwartz, L.H.: An automatic method for ground glass opacity nodule detection and segmentation from CT studies. In: *IEEE EMBS*, vol. 1, pp. 3062–3065 (2006)
5. Ye, X., Lin, X., Beddoe, G., Dehmeshki, J.: Efficient computer-aided detection of ground-glass opacity nodules in thoracic CT images. In: *Proceedings of the 29th Annual International Conference of the IEEE EMBS*. vol. 1, pp. 4449–4452 (2007)
6. Tao, Y., Lu, L., Dewan, M., Chen, A.Y., Corso, J., Xuan, J., Salganicoff, M., Krishnan, A.: Multi-level ground glass nodule detection and segmentation in CT lung images. In: Yang, G.Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) *MICCAI 2009*. LNCS, vol. 5762, pp. 715–723. Springer, Heidelberg (2009)
7. Heitmann, K.R., Kauczor, H., Mildenerberger, P., Uthmann, T., Perl, J., Thelen, M.: Automatic detection of ground glass opacities on lung HRCT using multiple neural networks. *Eur. Radiol.* 7(9), 1463–1472 (1997)
8. Kauczor, H.U., Heitmann, K., Heussel, C.P., Marwede, D., Uthmann, T., Thelen, M.: Automatic detection and quantification of ground-glass opacities on high-resolution CT using multiple neural networks: Comparison with a density mask. *AJR Am. J. Roentgenol.* 175(5), 1329–1334 (2000)
9. van Rikxoort, E.M., de Hoop, B., Viergever, M.A., Prokop, M., van Ginneken, B.: Automatic lung segmentation from thoracic computed tomography scans using a hybrid approach with error detection. *Med. Phys.* 36(7), 2934–2947 (2009)
10. van Ginneken, B., Baggerman, W., van Rikxoort, E.M.: Robust segmentation and anatomical labeling of the airway tree from thoracic CT scans. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) *MICCAI 2008, Part I*. LNCS, vol. 5241, pp. 219–226. Springer, Heidelberg (2008)
11. Godoy, M.C.B., Naidich, D.P.: Subsolid pulmonary nodules and the spectrum of peripheral adenocarcinomas of the lung: recommended interim guidelines for assessment and management. *Radiology* 253(3), 606–622 (2009)
12. Hu, M.K.: Visual pattern recognition by moment invariants, computer methods in image analysis. *IRE Transactions on Information Theory* 8, 179–187 (1962)
13. Ojala, T., Pietikainen, M., Maenpää, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* 24(7), 971–987 (2002)
14. Frangi, A.F., Niessen, W.J., Vincken, K.L., Viergever, M.A.: Multiscale vessel enhancement filtering. In: Wells, W., Colchester, A., Delp, S. (eds.) *MICCAI 1998*. LNCS, vol. 1496, pp. 130–137. Springer, Heidelberg (1998)
15. Friedman, J., Hastie, T., Tibshirani, R.: Additive logistic regression: A statistical view of boosting. *Ann. Stat.* 28(2), 337–407 (2000)
16. Jobson, J.D.: *Applied Multivariate Data Analysis*. Springer, Heidelberg (1992)

Learning Shape and Texture Characteristics of CT Tree-in-Bud Opacities for CAD Systems

Ulas Bagci¹, Jianhua Yao¹, Jesus Caban², Anthony F. Suffredini³,
Tara N. Palmore⁴, and Daniel J. Mollura¹

¹ Department of Radiology and Imaging Sciences,

² National Library of Medicine

³ Critical Care Medicine Department

⁴ Laboratory of Clinical Infectious Diseases, National Institutes of Health (NIH),
Bethesda, MD, USA
`ulas.bagci@nih.gov`

Abstract. Although radiologists can employ CAD systems to characterize malignancies, pulmonary fibrosis and other chronic diseases; the design of imaging techniques to quantify infectious diseases continue to lag behind. There exists a need to create more CAD systems capable of detecting and quantifying characteristic patterns often seen in respiratory tract infections such as influenza, bacterial pneumonia, or tuberculosis. One of such patterns is Tree-in-bud (TIB) which presents *thickened* bronchial structures surrounding by clusters of *micro-nodules*. Automatic detection of TIB patterns is a challenging task because of their weak boundary, noisy appearance, and small lesion size. In this paper, we present two novel methods for automatically detecting TIB patterns: (1) a fast localization of candidate patterns using information from local scale of the images, and (2) a Möbius invariant feature extraction method based on learned local shape and texture properties. A comparative evaluation of the proposed methods is presented with a dataset of 39 laboratory confirmed viral bronchiolitis human parainfluenza (HPIV) CTs and 21 normal lung CTs. Experimental results demonstrate that the proposed CAD system can achieve high detection rate with an overall accuracy of 90.96%.

Keywords: Tree-in-Bud, Willmore Energy, Lung, Infectious Diseases, Computer Assisted Detection.

1 Introduction

As shown by the recent pandemic of novel swine-origin H1N1 influenza, respiratory tract infections are a leading cause of disability and death. A common image pattern often associated with respiratory tract infections is TIB opacification, represented by thickened bronchial structures locally surrounded by clusters of 2-3 millimeter micro-nodules. Such patterns generally represent disease of the small airways such as infectious-inflammatory bronchiolitis as well as bronchiolar luminal impaction with mucus, pus, cells or fluid causing normally invisible peripheral airways to become visible [1]. Fig. 1 shows TIB patterns in a chest CT.

The precise quantification of the lung volume occupied by TIB patterns is a challenging task limited by significant inter-observer variance with inconsistent visual scoring methods. These limitations raise the possibility that radiologists' assessment of respiratory tract infections could be enhanced through the use of computer assisted detection (CAD) systems. However, there are many technical obstacles to detecting TIB patterns because micro-nodules and abnormal peripheral airway structures have strong shape and appearance similarities to TIB patterns and normal anatomic structures in the lungs.

In this work, we propose a new CAD system to evaluate and quantify respiratory tract infections by automatically detecting TIB patterns. The main contributions of the paper are two-fold: (1) A candidate selection method that locates possible abnormal patterns in the images. This process comes from a learning perspective such that the size, shape, and textural characteristics of TIB patterns are learned a priori. The candidate selection process removes large homogeneous regions from consideration which results in a fast localization of candidate TIB patterns. The local regions enclosing candidate TIB patterns are then used to extract shape and texture features for automatic detection; (2) another novel aspect in this work is to extract Möbius invariant local shape features. Extracted local shape features are combined with statistical texture features to classify lung tissues. To the best of our knowledge, this is the first study that uses automatic detection of TIB patterns for a CAD system in infectious lung diseases. Since there is no published work on automatic detection of TIB patterns in the literature, we compare our proposed CAD system on the basis of different feature sets previously shown to be successful in detecting lung diseases in general.

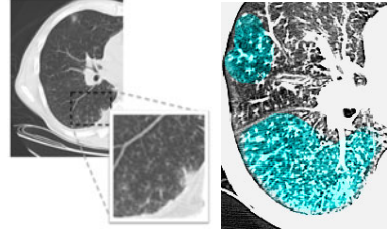


Fig. 1. (Left) CT image with a significant amount TIB patterns. (Right) Labelled TIB patterns (blue) in zoomed window on the right lung.

2 Methodology

The proposed CAD methodology is illustrated in Fig. 2. First, lungs are segmented from CT volumes. Second, we use locally adaptive scale based filtering method to detect candidate TIB patterns. Third, segmented lung is divided into local patches in which we extract invariant shape features and statistical texture features followed by support vector machine (SVM) classification. We extract features from local patches of the segmented lung only if there are candidate TIB patterns in the patches. The details of the proposed methods are presented below.

I. Segmentation. Segmentation is often the first step in CAD systems. There are many clinically accepted segmentation methods in clinics [23]. In this study,

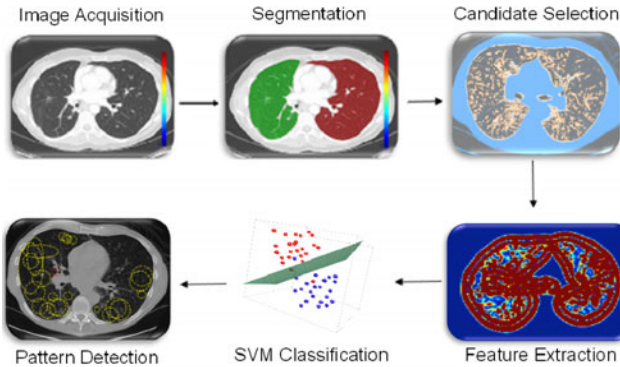


Fig. 2. The flowchart of the proposed CAD system for automatic TIB detection

fuzzy connectedness (FC) image segmentation algorithm is used to achieve successful delineation [2]. In FC framework, left and right lungs are “recognized” by automatically assigned seeds, which initiate FC segmentation.

II. Learning characteristics of TIB patterns. From Fig. 1, we can readily observe that TIB patterns have intensity characteristics with high variation towards nearby pixels, and such regions do not usually exceed a few millimetre (mm) in length. In other words, TIB patterns do not constitute sufficiently large homogeneous regions. Non-smooth changes in local gradient values support this observation. As guided by these observations, we conclude that (a) TIB patterns are *localized only in the vicinity of small homogeneous regions*, and (b) their boundaries have high curvatures due to the nature of its complex shape.

III. Candidate Pattern Selection. Our candidate detection method comes from a learning perspective such that we assign every internal voxel of the lung a membership value reflecting the size (i.e., scale) of the homogeneous region that the voxel belongs to. To do this, we use a locally adaptive scale based filtering method called ball-scale (or b-scale for short) [2]. b-scale is the *simplest* form of a locally adaptive scale where the scene is partitioned into several scale levels within which every voxel is assigned the size of the local structure it belongs. For instance, voxels within the large homogeneous objects have highest scale values, and the voxels nearby the boundary of objects have small scale values. Because of this fact and the fact in II.(a), we draw the conclusion that TIB patterns constitute only small b-scale values, hence, it is highly reasonable to consider voxels with small b-scale values as candidate TIB patterns. Moreover, it is indeed highly practical to discard voxels with high b-scale values from candidate selection procedure. Fig. 2 (candidate selection) and Fig. 3(b) show selected b-scale regions as candidate TIB patterns. A detailed description of the b-scale algorithm is presented in [2].

3 Feature Extraction

For a successful CAD system for infectious lung diseases, there is a need to have representative features characterizing shape and texture of TIB patterns efficiently. Since TIB is a complex shape pattern consisting of curvilinear structures with nodular structures nearby (i.e., a budding tree), we propose to use local shape features (derived from geometry of the local structures) combined with grey-level statistics (derived from a given local patch).

It has been long known that curvatures play an important role in the representation and recognition of intrinsic shapes. However, similarity of curvature values may not necessarily be equivalent to intrinsic shape similarities, which causes a degradation in recognition and matching performance. To overcome this difficulty, we propose to use Willmore energy functional [4] and several different affine invariant shape features parametrically related to the Willmore energy functional.

Willmore Energy. The Willmore energy of surfaces plays an important role in digital geometry, elastic membranes, and image processing. It is closely related to Canham-Helfrich model, where surface energy is defined as

$$\mathcal{S} = \int_{\Sigma} \alpha + \beta(H)^2 - \gamma K dA. \quad (1)$$

This model is curvature driven, invariant under the the group of Möbius transformations (in particular under rigid motions and scaling of the surface) and shown to be very useful in energy minimization problems. Invariance of the energy under rigid motions leads to conservation of linear and angular momenta, and invariance under scaling plays a role in setting the size of complex parts of the intrinsic shapes (i.e., corners, wrinkles, folds). In other words, the position, grey-level characteristics, size and orientation of the pattern of interest have minimal effect on the extracted features as long as the suitable patch is reserved for the analysis. In order to have simpler and more intuitive representation of the given model, we simply set $\alpha = 0$ and $\beta = \gamma = 1$, and the equation turns into the Willmore energy functional,

$$\mathcal{S}_w = \int_{\Sigma} (H^2 - K) dA = \int_{\Sigma} |H|^2 dA - \int_{\partial\Sigma} |K| ds, \quad (2)$$

where H is the mean curvature vector on Σ , K the Gaussian curvature on $\partial\Sigma$, and dA , ds the induced area and length metrics on Σ , $\partial\Sigma$ (representing area and boundary, respectively). Since homogeneity region that a typical TIB pattern appears is small in size, total curvature (or energy) of that region is high and can be used as a discriminative feature.

In addition to Willmore energy features, we have included seven different local shape features in the proposed CAD system. Let κ_1 and κ_2 indicate eigenvalues of the local Hessian matrix for any given local patch, the following shape features are extracted: 1) mean curvature (H), 2) Gaussian curvature (K),

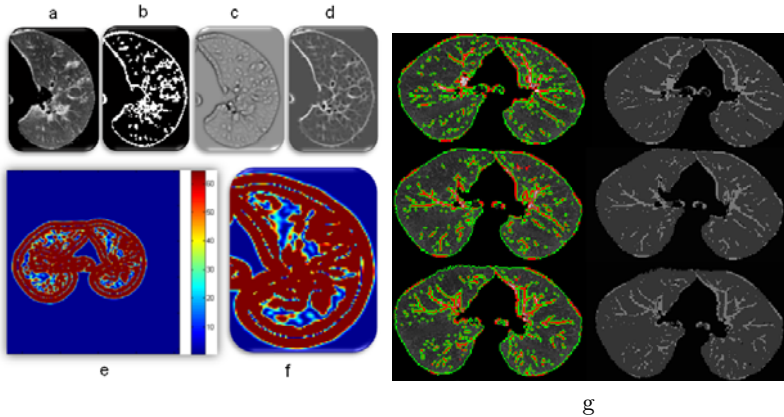


Fig. 3. a. CT lung, b. selected b-scale patterns, c. mean Curvature map (H), d. Gaussian Curvature (K), e. Willmore energy map, f. zoomed (e). g. Multi-phase level set segmentation based on the proposed shape features is shown in three different slices from the same patient’s chest CT scan.

3) shape index (SI), 4) elongation (κ_1/κ_2), 5) shear ($(\kappa_1 - \kappa_2)^2/4$), 6) compactness ($1/(\kappa_1\kappa_2)$), and 7) distortion ($\kappa_1 - \kappa_2$). Briefly, the SI is a statistical measure used to define local shape of the localized structure within the image [5]. Elongation indicates the *flatness* of the shape. Compactness feature measures the similarity between shape of interest and a perfect ellipse. Fig. 3(c) and (d) show mean and Gaussian curvature maps from which all the other local shape features are extracted. Fig. 3(e) and (f) show Willmore energy map extracted from Fig. 3(a).

Based on the observation in training, TIB patterns most likely occur in the regions inside the lung with certain ranges (i.e, blue and yellow regions). This observation facilitates one practically useful fact in the algorithm that, in the feature extraction process, *we only extract features if and only if at least “one” b-scale pattern exists in the local region as well as Willmore energy values of pixels lie in the interval observed from training.* Moreover, considering the Willmore energy has a role as hard control on feature selection and computation, it is natural to investigate their ability to segment images. We present a segmentation framework in which every voxel is described by the proposed shape features. A multi-phase level set [6] is then applied to the resulting vectorial image and the results are shown in Fig. 3(g). First and second columns of the Fig. 3(g) show segmented structures and the output homogeneity maps showing segmented regions in different grey-level, respectively. Although segmentation of small airway structures and pathological patterns is not the particular aim of this study, the proposed shape features show promising results due to their discriminative power.

Texture features. Spatial statistics based on Grey-Level Co-occurrence Matrix (GLCM) [7] are shown to be useful in discriminating patterns pertaining to lung diseases. As texture can give a lot of insights into the classification and characterization problem of poorly defined lesions, regions, and objects, we combine our proposed shape based invariants with GLCM based features. We extract 18 GLCM features from each local patch including autocorrelation, entropy, variance, homogeneity, and extended features of those. Apart from the proposed method, we also compare our proposed method with well known texture features: steerable wavelets (computed over 1 scale and 6 orientations with derivative of Gaussian kernel), GLCM, combination of shape and steerable wavelets, and considering different local patch size.

4 Experimental Results

39 laboratory confirmed CTs of HPIV infection and 21 normal lung CTs were collected for the experiments. The in-plane resolution is affected from patients' size and varying from 0.62mm to 0.82mm with slice thickness of 5mm. An expert radiologist carefully examined the complete scan and labeled the regions as normal and abnormal (with TIB patterns). As many regions as possible showing abnormal lung tissue were labeled (see Table 1 for details of the number of regions used in the experiments). After the proposed CAD system is tested via two-fold cross validations with labeled dataset, we present receiver operator characteristic (ROC) curves of the system performances.

Table 1 summarizes the performance of the proposed CAD system as compared to different feature sets. The performances are reported as the areas under the ROC curves (A_z). Note that shape features alone are superior to other

Table 1. Accuracy (A_z) of the CAD system with given feature sets

Features	Dimension	Patch Size	# of patches (TIB)	# of patches (Normal)	Area under ROC curve: A_z
Shape & GLCM	8+18=26	17x17	14144	12032	0.8991
Shape & GLCM	8+18=26	13x13	24184	20572	0.9038
Shape & GLCM	8+18=26	9x9	50456	42924	0.9096
Shape	8	17x17	14144	12032	0.7941
Shape	8	13x13	24184	20572	0.7742
Shape	8	9x9	50456	42924	0.7450
Steer. Wavelets& Shape	6x17x17+8=1742	17x17	14144	12032	0.7846
Steer. Wavelets& Shape	6x13x13+8=1022	13x13	24184	20572	0.7692
Steer. Wavelets& Shape	6x9x9+8=494	9x9	50456	42924	0.7908
Steer. Wavelets	6x17x17=1734	17x17	14144	12032	0.7571
Steer. Wavelets	6x13x13=1014	13x13	24184	20572	0.7298
Steer. Wavelets	6x9x9=486	9x9	50456	42924	0.7410
GLCM	18	17x17	14144	12032	0.7163
GLCM	18	13x13	24184	20572	0.7068
GLCM	18	9x9	50456	42924	0.6810

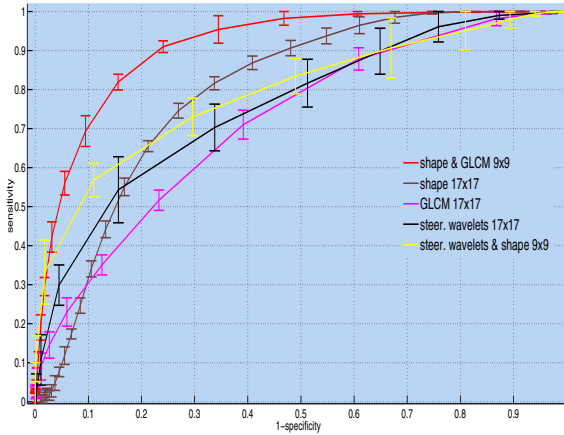


Fig. 4. Comparison of CAD performances via ROC curves of different feature sets

Table 2. p-values are shown in confusion matrix

p-Confusion Matrix	Shape	Steer.& Shape	Steer.	GLCM
Shape& GLCM	0.0171	0.0053	0.0056	0.0191
Shape	—	0.0086	0.0094	0.0185
Steer.& Shape	—	—	0.0096	0.0175
Steer.	—	—	—	0.0195

methods even though the dimension of the shape feature is only 8. The best performance is obtained when we combined shape and GLCM features. This is expected because spatial statistics are incorporated into the shape features such that texture and shape features are often complementary to each other. In what follows, we select the best window size for each feature set and plot their ROC curves all in Fig. 4. To have a valid comparison, we repeat candidate selection step for all the methods, hence, the CAD performances of compared feature sets might perhaps have lower accuracies if the candidate selection part is not applied. Superiority of the proposed features is clear in all cases. To show whether the proposed method is significantly different than the other methods, we compared the performances through paired t-tests, and the p-values of the tests are summarized in Table 2. Note that statistically significant changes are emphasized by $p < .01$ and $p < .05$.

5 Conclusion

In this paper, we have proposed a novel CAD system for automatic TIB pattern detection from lung CTs. The proposed system integrates 1) fast localization of candidate TIB patterns through b-scale filtering and scale selection, and 2) combined shape and textural features to identify TIB patterns. Our proposed shape features illustrate the usefulness of the invariant features, Willmore energy features in particular, to analyze TIB patterns in Chest CT. In this paper, we have not addressed the issue of quantitative evaluation of severity of diseases by expert observers. This is a challenging task for complex shape patterns such as TIB opacities, and subject to further investigation.

References

1. Eisenhuber, E.: The tree-in-bud sign. *Radiology* 222(3), 771–772 (2002)
2. Saha, P.K., Udupa, J.K., Odhner, D.: Scale-based fuzzy connected image segmentation: Theory, algorithms, and validation. *Computer Vision Image Understanding* 77, 145–174 (2000)
3. Hu, S., Hoffman, E.A., Reinhardt, J.M.: Automatic lung segmentation for accurate quantification of volumetric X-ray CT images. *Transactions on Medical Imaging* 20(6), 490–498 (2001)
4. Bobenko, E.I., Shroder, P.: Discrete Willmore Flow. In: *Eurographics Symposium on Geometry Processing*, pp. 101–110 (2005)
5. Lillholm, M., Griffin, L.D.: Statistics and category systems for the shape index descriptor of local 2nd order natural image structure. *Image and Vision Computing* 27(6), 771–781 (2009)
6. Ayed, I.B., Mitiche, A., Belhadj, Z.: Polarimetric image segmentation via maximum-likelihood approximation and efficient multiphase level-sets. *Transactions on Pattern Analysis and Machine Intelligence* 28(9), 1493–1500 (2006)
7. Haralick, R.M., Shanmugam, K., Dinstein, I.: Textural Features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics* 3(6), 610–621 (1973)

Referenceless Stratification of Parenchymal Lung Abnormalities

Sushravya Raghunath, Srinivasan Rajagopalan, Ronald A. Karwoski,
Brian J. Bartholmai, and Richard A. Robb

Mayo Clinic College of Medicine, Rochester, MN, USA
{raghunath.sushravya,rajagopalan.srinivasan,karwoski.ronald,
bartholmai.brian,robb.richard}@mayo.edu

Abstract. This paper introduces computational tools that could enable personalized, predictive, preemptive, and participatory (P4) Pulmonary medicine. We demonstrate approaches to (a) stratify lungs from different subjects based on the spatial distribution of parenchymal abnormality and (b) visualize the stratification through glyphs that convey both the grouping efficacy and an iconic overview of an individual's lung wellness. Affinity propagation based on regional parenchymal abnormalities is used in the referenceless stratification. Abnormalities are computed using supervised classification based on Earth Mover's distance. Twenty natural clusters were detected from 372 CT lung scans. The computed clusters correlated with clinical consensus of 9 disease types. The quality of inter- and intra-cluster stratification as assessed by ANOSIM R was 0.887 ± 0.18 ($pval < 0.0005$). The proposed tools could serve as biomarkers to objectively diagnose pathology, track progression and assess pharmacologic response within and across patients.

Keywords: Referenceless stratification, affinity propagation, idiopathic pulmonary fibrosis, glyphs.

1 Introduction

The disease processes in Diffuse Parenchymal Lung Disease (DPLD) are characterized by distinct cellular infiltrates and extracellular matrix deposition and are broadly classified in CT scans into five primal forms- normal, emphysema, ground glass, honeycombing and reticular. The distribution of these patterns through the lung lobes is indicative of a specific DPLD disease. Differentiability of different DPLDs is central to early application of appropriate therapy so as to positively affect patient prognosis. Existing clinical decision support tools are suboptimal for consistent characterization and visual representation of the type and extent of disease in a clinician and patient friendly manner.

Image based stratification of parenchymal abnormalities specific to patient data in a referenceless manner (with no alignment to a common coordinate system) would facilitate consistent assessment of patient lung wellness. Previously, referenceless unbiased stratification has been proposed for cardiac images [12].

No such approach exists for the lung, especially to stratify the biological behavior of abnormalities across patient populations. We address this unmet need using a computational tool to stratify and a visualization tool to provide an iconic overview of the stratifications.

Lung abnormalities were determined in terms of the primal CT appearance forms through supervised classification. During the training phase, radiologists selected multiple volumes of interest (VOIs) to represent the training sets for each of the primal forms. For the supervised classification, earth mover's distance was used as the similarity metric between the histograms of a voxel's neighborhood and the labeled training set. The spatial distribution of the primal forms across the lobes was computed and mapped into glyphs. A new pairwise dissimilarity metric, *spikelets*, and affinity propagation [3] were used to stratify abnormality distribution across the lung lobes. The clusters were analyzed statistically with Analysis of Similarity (ANOSIM). The exemplars and their respective candidates were automatically categorized into clinical groups using the guidelines in [4]. The glyphs and the underlying classifications were verified by the radiologist to assess correctness. Our experimental results suggest that the proposed tool is a valuable technology to realize the potential of P4 medicine through direct improvement of quality and consistency in clinical pulmonary practice.

2 Materials and Methods

2.1 Datasets, VOIs and Classification

CT scans from 372 patients with disease across the DPLD spectrum were used for this study. The patients were scanned on a HRCT scanner (140 kVp, 250 mAs, BONE kernel recon, 1.25mm slice thickness, 50% overlap with 0.625 mm³ voxels). Three radiologists screened the scans and selected a subset of 14 datasets to pick multiple 15³ VOIs across the primal forms. 976 VOIs were selected to represent 80, 150, 187, 265, and 294 VOIs of emphysema, ground glass, honey combing, normal and reticular forms, respectively. The lungs, airways and vessels were extracted using readily available techniques. During supervised nearest neighbor classification, the histogram of 15³ neighborhood around each parenchymal voxel was compared against the VOI histograms using Earth Mover's Distance.

2.2 Regional Abnormality Distribution and Glyph Creation

The lobar regions of the individual lungs were segmented using an approach similar to that described in [5]. The distribution of the five primal forms within the respective lobes was computed and used to create the glyphs. Figure 1 shows a representative glyph. The glyph is divided into six regions each representing a lung lobe. The lobes are uniquely labeled to indicate their spatial location. The first letter (R/L) denotes the right and left lung, the second letter (U/M/L) denotes respectively the upper, middle and lower lobes. Although no pleural separation demarcates the lingula from the remainder of the LU lobe, this anatomic region was defined as LM lobe by reflection from the right lung. The origin of

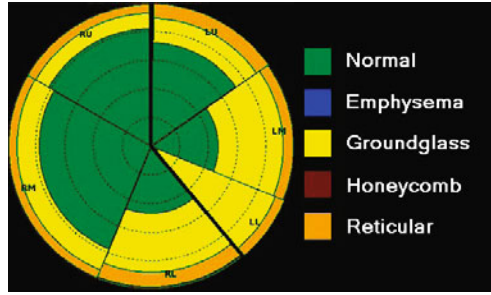


Fig. 1. Glyph layout of the distribution of abnormal patterns across different lobes of the lung

the glyph is fixed at 12-o'clock starting with the LU lobe. The asymmetry, if any, between the left and right lungs can be readily observed in the glyph. The individual lobes span through angles proportional to their respective volumes. Within each lobe, the disease distribution is represented by color-coded sectors proportional to the percentage of disease in that lobe. The concentric circles are drawn at 20% intervals for enhanced visualization.

To create an unbiased stratification of the abnormality distribution, affinity propagation [3] - an unsupervised clustering technique that automatically finds the natural number of clusters- was used. Affinity propagation uses message passing to iteratively find clusters from pair-wise dissimilarities of n -dimensional data. In addition to resolving the clusters, it identifies the exemplar that is most 'central' to each of the clusters. Pairwise dissimilarity between given lungs A and B based on their respective lobar distribution of the five primal patterns was computed as

$$D(A, B) = \sum_{R=1:6} \alpha_R * d_R(A, B); \quad d_R(A, B) = \sum_{i=1:5} spikelet(A_i^R, B_i^R) \quad (1)$$

$$spikelet(A_i^R, B_i^R) = \begin{cases} \frac{|A_i^R - B_i^R|}{\max(A_i^R, B_i^R)} & \text{for } \max(A_i^R, B_i^R) > acuity \\ 0 & \text{otherwise,} \end{cases} \quad (2)$$

where, $\alpha_R = 100/\overline{Vol}_R$; \overline{Vol}_R is the mean lobar volume; A_i^R is the % distribution of i^{th} primal pattern in the R^{th} lobe of lung A ; and $acuity$ is the minimum resolvable distribution differences, which was set to 2% in our experiments.

The clusters were categorized automatically into clinical groups using the guidelines suggested in [4]. Accordingly, based on the type and extent of abnormality distribution of the individual exemplar, its cluster was labeled as one of the following 9 types: (T1) diffuse emphysema (T2) upper-lobe emphysema (T3) emphysema with early fibrosis (T4) probable NSIP (non-specific interstitial pneumonitis), (T5) confident NSIP, (T6) NSIP with concurrent emphysema, (T7) probable UIP (usual interstitial pneumonitis), (T8) confident UIP and (T9) UIP with concurrent emphysema. The quality of clusters was evaluated by radiological review and statistical analysis (ANOSIM).

3 Results

Figure 2 shows representative results of the supervised classification. The voxels in panel B are color-coded using the scheme given in Figure 1. Panel C shows the 3D rendering of the lobe definitions. The classifications were visually verified by a radiologist. The classifications and the lobe definitions of the individual lungs were used to compute the primal pattern distribution and subsequently the individual glyphs. Figure 3 shows a mosaic of the glyphs from all 372 CT scans. Accounting for the voxel sizes and the number of voxels classified, it is worth noting that information collated across a 1.86 meter³ space has been captured in the 5x5 inch mosaic space. Even at this resolution, the glyphs provide a succinct overview of the entire database and highlight the ease with which the intra- and inter-patient disease distribution can be pre-attentively captured. With effortless effectiveness it is easy to visualize the presence of distinct cases with Chronic Obstructive Pulmonary Disease (COPD- emphysematous glyphs with blue shades), NSIP- variable amounts of ground glass and reticular distributions and UIP- extensive reticulation and honeycombing.

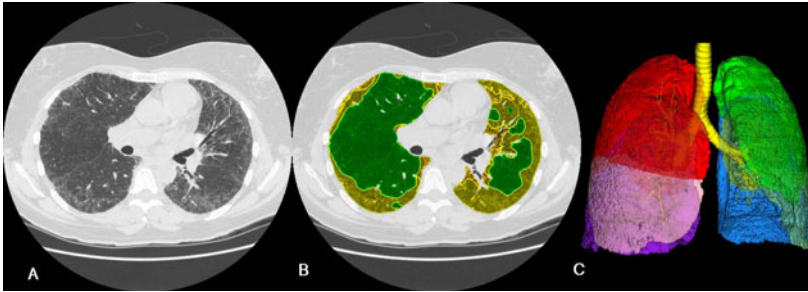


Fig. 2. Representative results of parenchymal tissue classification and 3D rendering of lobar distribution

Figure 4 shows qualitative results of the referenceless stratification. Panels A and B show the pair-wise similarities before and after stratification; the darker shade implies low similarity between two datasets. Affinity propagation on the original matrix (panel A) yielded 20 unique clusters. The 20 clusters correspond to the 20 diagonal sub-blocks shown (by red boxes along the diagonal) in panel B; the stratification qualitatively reveals the maximization and minimization, respectively, of intra and inter (off-diagonal sub-blocks) cluster similarity. The maximum, minimum and mean number of candidates in the clusters were 54, 8 and 19, respectively.

Quantitative efficacy of the stratification was examined using ANOSIM R to assess the magnitude of the differences among clusters. An R value of 1 suggests that the communities completely differ among the defined groups, a value of 0 indicates no difference among groups. The pair-wise inter-cluster R values shown

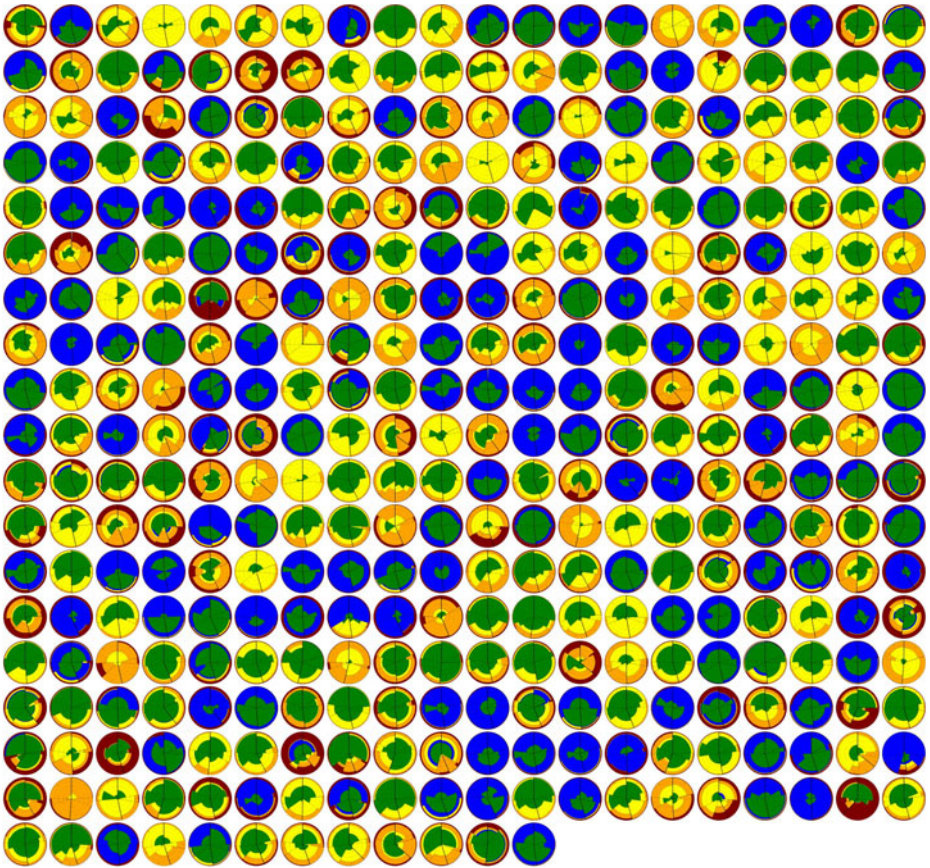


Fig. 3. Mosaic of glyphs for all 372 datasets used in this study. The glyphs provide a succinct overview of inter- and intra-subject distribution of parenchymal abnormalities.

Figure 5A highlights the optimality of the stratification. The combined R for the 20 clusters was 0.887 ± 0.18 ($pval < 0.0005$) highlighting the greater agreement of the candidates within each of the clusters. Figure 5B shows the mean intra-cluster, exemplar-global and inter-exemplar distances for six representative clusters. The tightness of intra-cluster distances quantitatively validates the visual representations of stratification efficacy.

The 20 clusters were automatically categorized into clinical groups based on the abnormality distribution of their respective exemplars. Figure 6 shows the glyphs for all the exemplars, along with their clinical categories. Differentiation of the individual emphysematous lungs into upper lobe predominance (T2) has profound effect on disease management due to better outcomes after lung volume resection surgery. Though the clinical guidelines suggest nine categories, the glyphs reveal the significant pathological variations within the categories. For example, confident UIP (T8), which has poor prognosis, has five distinct

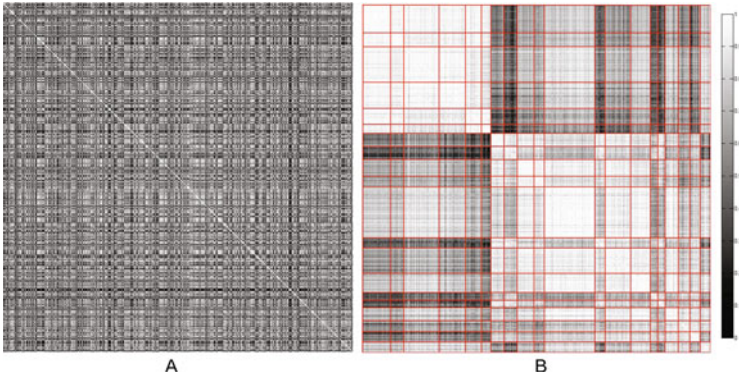


Fig. 4. Pairwise similarity between lungs before (A) and after (B) affinity propagation based clustering. Lighter the shade; higher the similarity.

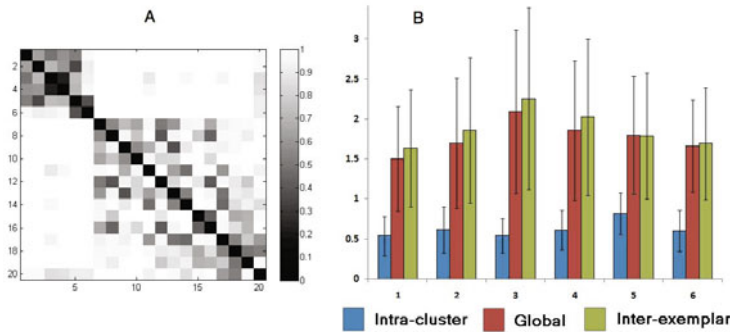


Fig. 5. Quantitative analysis of stratification showing intra-cluster pair-wise ANOSIM R values (A) and comparison of intra-cluster tightness with inter-exemplar and global similarities for representative clusters (B)

variations in the regional distribution and extent of reticular and honeycombing patterns. Lumping such distinct variations into a single category could prevent the delivery of personalized and expeditious clinical care. The diagnostic disparity in performance between physicians based in academic versus community centers is well known. Such disparity is disturbing and could (at times, irreversibly) compromise patient care. The quantitative stratification proposed in this paper shows promise to reduce this discrepancy.

Figure 7 shows representative glyphs across the spectrum of diffuse pulmonary lung diseases. While the confident categories of UIP (T8) and NSIP (T5) are easily differentiable by manual radiological reviews, even after accommodating the errors due to subjective aggregation, probable categories of UIP (T7) and NSIP (T4) are often misinterpreted. The upper lobe dominance of emphysema (T2) is usually assessed using a count of voxels beyond a certain threshold. Such quantification is extremely sensitive to image slice thickness, acquisition parameters,

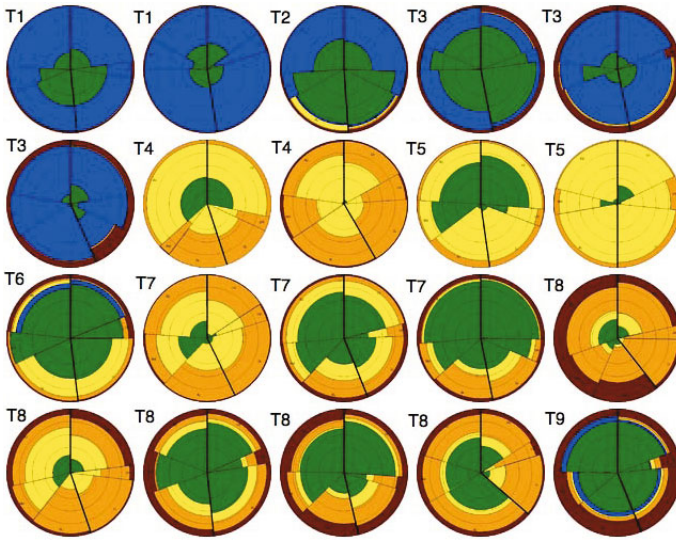


Fig. 6. The glyphs for the 20 cluster exemplars along with their categorization into clinical groups

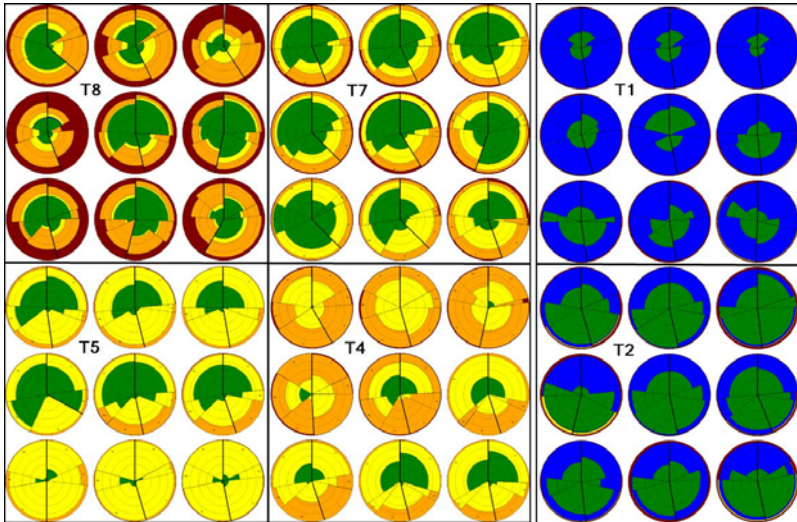


Fig. 7. Representative glyphs for the natural clusters categorized as confident UIP (T8), probable UIP (T7), Diffuse Emphysema (T1), confident NSIP (T5), probable NSIP (T4) and Upper-lobe Emphysema (T2)

and the reconstruction kernel utilized. On the other hand, the quantitative stratification and glyphs described here provide unambiguous categorization of the disease state.

4 Discussion and Conclusion

Current management of diffuse parenchymal lung disease might best be described as a random walk through therapy space. However, recent advances in imaging offer opportunities to develop and validate lung-specific biomarkers as potential cross correlates to diagnosis, staging of treatment and therapy monitoring. Despite the enhanced contrast and spatial resolution of HRCT scans, classification and quantification of interstitial lung disease is difficult, and even experienced chest radiologists are challenged with differential diagnosis. Robust, expeditious and reproducible segmentation and characterization of the lung, lobes, airways, vessels and parenchymal tissues, accompanied by results summarized holistically and presented in a consistent manner, will advance the field of Computer Aided Diagnosis and elevate it to a degree of maturity and universal applicability heretofore not obtained. The stratification and visualization strategy proposed in this paper represents a major step towards harnessing the power of information technology and image computing to develop a computational framework that enables a powerful and hitherto elusive capability for pulmonary imaging evaluation: a trustable, verifiable, and clinically relevant comprehensive summary of pulmonary disease, including the extent and character of disease, both within an individual patient and across a cohort of patients.

References

1. De Craene, M., du Bois d'Aische, A., Macq, B., Warfield, S.K.: Multi-subject Registration for Unbiased Statistical Atlas Construction. In: Barillot, C., Haynor, D.R., Hellier, P. (eds.) MICCAI 2004. LNCS, vol. 3216, pp. 655–662. Springer, Heidelberg (2004)
2. Srinivasan, R., Shriram, K.S., Suryanarayanan, S.: Unbiased Stratification of Left Ventricles. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 551–558. Springer, Heidelberg (2008)
3. Frey, B., Dueck, D.: Clustering by Passing Messages Between Data Points. *Science* 315, 972–976 (2007)
4. Akira, M., Inoue, Y., Kitaichi, M., Yamamoto, S., Arai, T., Toyokawa, K.: Usual Interstitial Pneumonia and Nonspecific Interstitial Pneumonia with and without concurrent emphysema. *Radiology* 251, 271–279 (2009)
5. Ross, J.C., Estépar, R.S.J., Díaz, A., Westin, C.-F., Kikinis, R., Silverman, E.K., Washko, G.R.: Lung Extraction, Lobe Segmentation and Hierarchical Region Assessment for Quantitative Analysis on High Resolution Computed Tomography Images. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5762, pp. 690–698. Springer, Heidelberg (2009)

Lung Texture Classification Using Locally–Oriented Riesz Components

Adrien Depeursinge^{1,2}, Antonio Foncubierta–Rodríguez¹,
Dimitri Van de Ville^{2,3}, and Henning Müller^{1,2}

¹ University of Applied Sciences Western Switzerland (HES–SO)

² University and University Hospitals of Geneva (HUG), Switzerland

³ Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland
`adrien.depeursinge@hevs.ch`

Abstract. We develop a texture analysis framework to assist radiologists in interpreting high–resolution computed tomography (HRCT) images of the lungs of patients affected with interstitial lung diseases (ILD). Novel texture descriptors based on the Riesz transform are proposed to analyze lung texture without any assumption on prevailing scales and orientations. A global classification accuracy of 78.3% among five lung tissue types is achieved using locally–oriented Riesz components. Comparative performance analysis with features derived from optimized grey–level co–occurrence matrices showed an absolute gain of 6.1% in classification accuracy. The adaptability of the Riesz features is demonstrated by reconstructing templates according to the first principal components of the lung textures. The balanced performance achieved among the various lung textures suggest that the proposed methods can complement human observers in HRCT interpretation, and opens interesting perspectives for future research.

Keywords: Texture analysis, Riesz transform, interstitial lung diseases, high–resolution computed tomography, computer–aided diagnosis.

1 Introduction

Successful diagnostic interpretation of medical images relies on two distinct processes. First, abnormal image patterns are identified (e.g., fibrous tissue, architectural distortion, ...) and, second, links between the patterns and possible diagnoses can be established [1]. Whereas the latter requires a deep understanding and comprehensive experience of the involved diseases, the former is closely related to visual perception. Interestingly, a large scale study on malpractice in radiology showed that the majority of errors in medical image interpretation are caused by perceptual misapprehensions [2]. Texture analysis is central to human image understanding and plays an important role in efficient characterization of biomedical tissue, that cannot be described in terms of shape or morphology [3]. As a consequence, computer–aided diagnosis (CAD) based on texture quantification in radiological images has been an active research field over the

past 20 years with the aim of reducing omission errors of pathological tissue by providing systematic second opinions to radiologists.

Texture analysis is the cornerstone for differentiating between pathological and healthy lung tissue of patients affected by interstitial lung diseases (ILD) in high-resolution computed tomography (HRCT). ILDs group more than 150 disorders of the lung tissue of varying origin and can be differentiated only by detecting subtle texture changes of the lung parenchyma with a characteristic distribution within the lung anatomy [4]. Interpreting HRCT images of the chest represents a challenge even for trained radiologists and lung specialists. Several studies have been conducted on the use of computerized lung texture classification to assist the radiologists in HRCT interpretation for ILDs or chronic obstructive pulmonary disease (COPD) starting from 1997 [5,6,7,8]. The success of the CAD system is intimately related to the ability of the visual features to catch and learn the subtle texture signatures specific to each lung tissue type, which are typically non-deterministic. Therefore, statistical approaches that are able to capture texture properties at any location, scale and orientation (i.e., affine-covariant) are required to achieve high tissue classification performance to complement human observers. Whereas more than 60 papers using texture analysis to classify lung tissue can be found in the literature of the past 15 years [9], research contributions on novel texture descriptors are still required as several papers [5,6,10,11] rely on texture features derived from grey-level co-occurrence matrices (GLCM) [12], oriented filters from Gaussian derivatives [7] or local binary patterns (LBP) [8]. The performance of these methods depends on the arbitrary choice of scales and/or orientations as well as a necessary grey-level reduction for GLCMs, the latter entailing the risk of losing precious information contained in the full bit depth of the original image. Wavelets and filtering techniques have the advantage of providing continuous responses when compared to the binary or categorical outputs of GLCMs or LBPs, which allows for a finer detection and quantification of transients in medical images and were successfully used for lung texture classification in [7,13,14]. Specific wavelet transforms yield multiscale, multi-orientation with infinitesimal angular precision (i.e., steerable filterbanks) and translation invariant (i.e., undecimated transforms) analysis, which allows to characterize textures without making a priori choices on the affine parameters [15].

In previous work [13,9] we used isotropic wavelet frames enabling texture analysis with translation and scale covariance as well as rotation invariance. The use of isotropic analysis was based on the assumption that no prevailing orientations are contained in the lung tissue patterns of 2D axial slices in HRCT. Three research contributions are proposed in this article. First, a novel texture characterization approach based on the Riesz transform yielding translation, scale and rotation covariance is introduced. Second, the assumption that lung tissue patterns are locally rotation-invariant is investigated by aligning textures using the local prevailing orientation. Third, principal component analysis (PCA) of the Riesz features is used to obtain templates that are discriminative for lung textures. The approaches are evaluated and compared using a dataset of 85

ILD cases with a large variety of lung tissue types and a realistic validation scheme based on a leave-one-patient-out (LOPO) cross-validation (CV). We used 13808 overlapping blocks from 2037 manually drawn regions of interest in 1225 2D HRCT slices to validate the proposed methods. A quantitative performance comparison with optimized GLCMs is carried out.

2 Material and Methods

2.1 Dataset

A database of 85 ILD cases containing HRCT image series with a slice thickness of 1mm, inter-slice distance of 10mm and hand-drawn regions annotated in a collaborative fashion by two radiologists with 15 and 20 years of experience at the University Hospitals of Geneva (HUG) [16] is used to evaluate the performance of the proposed approaches. The diagnosis of each case was confirmed either by pathology (biopsy, bronchoalveolar washing) or by a laboratory/specific test. Based on [4], the texture classes are defined as healthy and four pathological lung tissue types (i.e., ground glass, fibrosis, micronodules and emphysema) that are used to characterize the most frequent ILDs in HRCT. The distribution of the annotated regions and patients is detailed in Table 1.

2.2 Texture Analysis with N th-Order Riesz Transforms

The Riesz transform is a multidimensional extension of the Hilbert transform, which maps any function $f(x)$ to its harmonic conjugate and is a very powerful tool for mathematical manipulations of periodic signals [17]. For a two-dimensional signal $f(\mathbf{x})$, the different components of the N th-order Riesz transform \mathcal{R} are defined in the Fourier domain as

$$\widehat{\mathcal{R}^{(n_1, n_2)} f(\boldsymbol{\omega})} = \sqrt{\frac{n_1 + n_2}{n_1! n_2!}} \frac{(-j\omega_1)^{n_1} (-j\omega_2)^{n_2}}{\|\boldsymbol{\omega}\|^{n_1 + n_2}} \hat{f}(\boldsymbol{\omega}), \quad (1)$$

for all combinations of (n_1, n_2) with $n_1 + n_2 = N$ and $n_{1,2} \in \mathbb{N}$. $\hat{f}(\boldsymbol{\omega})$ denotes the Fourier transform of $f(\mathbf{x})$, where the vector $\boldsymbol{\omega}$ is composed by $\omega_{1,2}$ corresponding to the frequencies in the two image axes. The multiplication by $j\omega_{1,2}$ in the numerator corresponds to partial derivatives of f and the division by the norm of $\boldsymbol{\omega}$ in the denominator results in only phase information being retained. Therefore, the 1st-order \mathcal{R} corresponds to an allpass filterbank with directional (singular) kernels $h_{1,2}$:

$$\mathcal{R}f(\mathbf{x}) = \begin{pmatrix} \mathcal{R}^{1,0} \\ \mathcal{R}^{0,1} \end{pmatrix} = \begin{pmatrix} h_1(\mathbf{x}) * f(\mathbf{x}) \\ h_2(\mathbf{x}) * f(\mathbf{x}) \end{pmatrix}, \quad (2)$$

where

$$h_{1,2}(\mathbf{x}) = \frac{x_{1,2}}{2\pi\|\mathbf{x}\|^3}, \quad (3)$$

and $x_{1,2}$ correspond to the axes of the image [15]. The Riesz transform commutes with translation, scaling or rotation. The orientation of the Riesz components

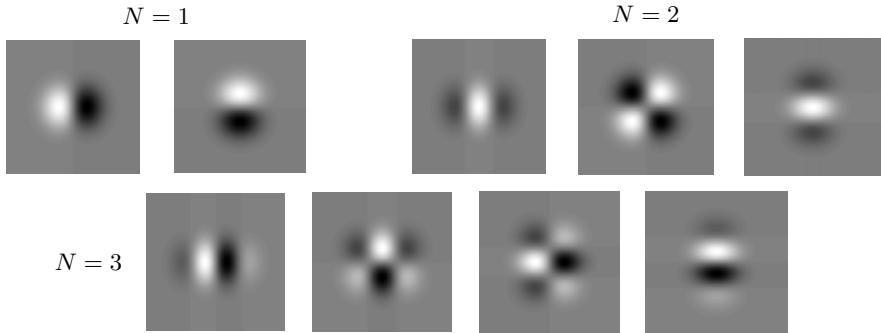


Fig. 1. Templates corresponding to the Riesz kernels convolved with a Gaussian smoother for $N=1,2,3$

is determined by the partial derivatives in Eq. (11). Whereas 2^N Riesz filters are generated by (11), only $N + 1$ components have distinct properties due to commutativity of the convolution operators in (2) (e.g., $\partial^2/\partial x\partial y$ is equivalent to $\partial^2/\partial y\partial x$). The Riesz components yield a steerable filterbank [15] allowing to analyze textures in any direction, which is an advantage when compared to classical Gaussian derivatives or Gabor filters. Qualitatively, the first Riesz component of even order corresponds to a ridge profile whereas for odd ones we obtain an edge profile, but much richer profiles can be obtained by linear combinations of the different components. The templates of $h_{1,2}(\mathbf{x})$ convolved with Gaussian kernels for $N=1,2,3$ are depicted in Fig. 1. The N th-order Riesz transform can be coupled with an isotropic multiresolution decomposition (e.g., Laplacian of Gaussian (LoG)) to obtain rotation-covariant (steerable) basis functions [15].

The main idea of the proposed approach is to derive texture signatures from multiscale Riesz coefficients. An example showing healthy and fibrosis tissue represented in terms of their Riesz components with $N=2$ is depicted in Fig. 2a). In order to provide a local categorization of the lung parenchyma, lung regions in 2D axial slices are divided into 32×32 overlapping blocks with a distance between contiguous block centers of 16. The Riesz transform is applied to each block, and every Riesz component $n = 1, \dots, N+1$ is mapped to a multiscale representation by convolving them with four LoG filters of scales $s = 1, \dots, 4$ with a dyadic scale progression. In a total of $(N+1) \times 4$ subbands, the variances $\sigma_{n,s}$ of the coefficients are used as texture features along with 22 grey level histogram (GLH) bins in $[-1050;600]$ Hounsfield Units (HU). The percentage of air pixels with values ≤ -1000 HU completes the feature space learned by support vector machines (SVM) with a Gaussian kernel.

The local dominant texture orientations have an influence on the repartition of respective responses of the Riesz components, which is not desirable for creating robust features with well-defined clusters of instances. For example, a rotation of $\pi/2$ will switch the responses of h_1 and h_2 for $N=1$. To ensure that the repartitions of $\sigma_{n,s}$ are comparable for two similar textures having distinct local

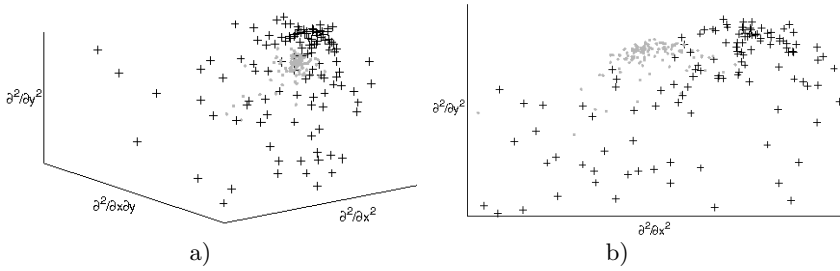


Fig. 2. Riesz representation of healthy (gray dots) versus fibrosis (black crosses) patterns. a) Initial Riesz coefficients in 3D. b) The Riesz coefficients in 2D after having locally aligned the texture based on local prevailing orientation. The component corresponding to $\partial^2/\partial x\partial y$ is zero after local rotation and is not shown in b).

prevailing orientations, the filters are oriented to have maximal response along h_1 . The dominant orientation θ_{dom} of h_1 at the position \mathbf{x}_p is

$$\theta_{dom}(\mathbf{x}_p) = \arg \max_{\theta \in [0, \pi]} \left(\left(h_1^{(\theta)} * g \right) * f \right) (\mathbf{x}_p), \quad (4)$$

where $h_1^{(\theta)}(\mathbf{x})$ is h_1 rotated by θ and $g(\mathbf{x})$ is a Gaussian kernel. A local orientation is obtained by rotating every Riesz filter h_n with θ_{dom} and is done analytically [15]. 2nd-order Riesz coefficients of healthy and fibrosis tissue after local orientation are shown in Fig. 2 b).

3 Results

The proposed methods are evaluated using blocks from annotated ROIs with a LOPO CV of 85 patients. For each fold, the cost C of the SVMs and the width σ_k of the Gaussian kernel are optimized with the training set where parameters allowing best classification accuracy on the training set are found with a grid search ($C \in [0.1, 1000]$, $\sigma_k \in [10^{-2}, 10^1]$) and a 5-fold CV. The classification performances are compared with optimized GLCMs that are extensively used for lung texture analysis in the literature. Texture features derived from GLCMs are contrast, correlation, energy and homogeneity for various pixel distances $d = 1, \dots, 5$ and orientations $\theta = 0, \pi/4, \pi/2, 3\pi/4$, similarly to [8]. Three grey-level reductions are compared: 8, 16 and 32 levels l . Optimized SVMs learn in the feature space spanned by concatenated GLCM attributes from every spacing and orientation parameters as well as GLH and air percentage. Classification accuracies using Riesz features of various orders ($N = 1, \dots, 13$) are compared before and after local rotation in Fig. 3 a). A class-specific performance comparison of best setups for Riesz, Riesz with local orientation, and GLCMs is shown in Fig. 3 b). The confusion matrix of the best performing technique ($N=6$ with local orientation) is detailed in Table 1. In Fig. 4, the distributions of the classes in

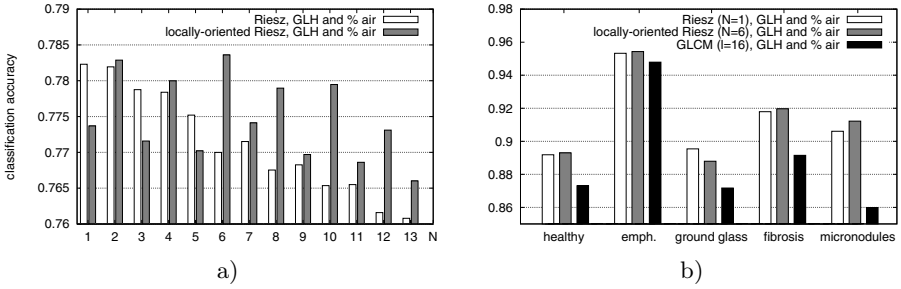


Fig. 3. a) Global classification accuracies with $N = 1, \dots, 13$. $N=6$ with local orientation reaches best performance with 78.3% correct predictions of 13808 instances. b) Class-specific accuracies of the two best configurations ($N=1$ and $N=6$ with orientation) and best performance of GLCMs ($l=16$).

Table 1. Confusion matrix in % of the best performing setup ($N=6$ with local orientation). The numbers of blocks and patients used for the evaluation are detailed. Note that a patient may have several types of lung tissue disorders.

	healthy	emphysema	ground glass	fibrosis	micronodules	# blocks	# patients
healthy	77.5	7.6	4.1	0	10.7	1975	7
emphysema	8.4	73.3	5.9	6.2	6.2	1298	6
ground glass	14.1	0.5	72.3	10	3.1	3513	32
fibrosis	0.7	2.6	8.4	84.5	3.8	3554	37
micronodules	11.6	0.7	3.5	3.7	80.5	3468	16

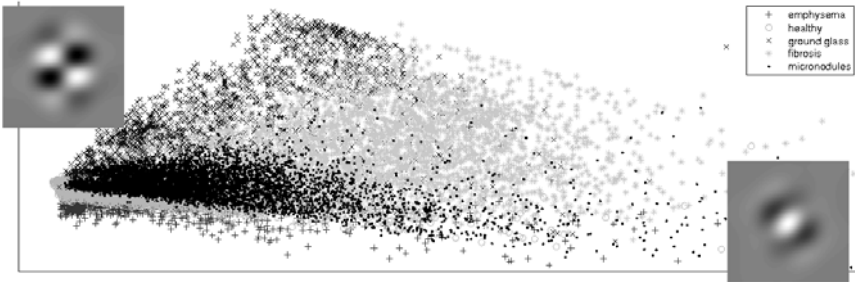


Fig. 4. Visualization of the feature space projected on the two dominant principal components. The corresponding Riesz templates (scale 2) are shown on the axes.

terms of the two dominant principal components of the 6th-order Riesz features with local orientation are shown. The coefficients from the PCA components are used to weight each Riesz component and create learned templates represented on axes in Fig 4.

4 Discussions and Conclusions

We propose a novel texture classification method based on the Riesz transform to categorize lung tissue patterns in HRCT image series from patients affected with ILDs. Compared to the literature, the Riesz features allow to analyze lung texture without prior assumptions on the prevailing scales and orientations and show higher classification performance than popular GLCM texture features. The inherent local orientations of the lung texture are studied by locally steering the Riesz components before further classification. An optimal order of $N=6$ with local orientation allows best classification performance with 78.3% correct predictions. Fig. 3 a) shows that even orders of the Riesz transform are providing best results, which suggests that ridge detectors are more appropriate than edge detectors for lung texture characterization. Fig. 3 b) shows that all lung tissue types except ground glass are better classified with local orientation. All lung tissue types benefit from Riesz texture features when compared to GLCM features. Classes with highest improvement are healthy, fibrosis and micronodules, which are those containing most texture information. The performance comparisons are statistically significant for all classes ($p \leq 0.0002$) but emphysema ($p = 0.073$). An absolute gain of 6.1% in global classification accuracy is obtained with Riesz and rotation adjustment (78.3%) when compared to GLCM (72.2%). This suggests that the arbitrary choices of scale and orientation parameters are not optimal for accurate characterization of the lung texture, although these values are the most commonly used in the literature [5,6,7,10,11,8]. Table 1 shows that healthy and ground glass patterns are the most challenging to separate due to high intra-class variability among patients and severity of disease. Confusion between micronodules and healthy tissue is observed, which is a limitation of 2D approaches as bronchovascular structures have similar appearance as micronodules in the 2D axial slices. Unfortunately, the HRCT imaging protocol is very anisotropic with a gap between slices of 10mm and does not allow for full 3D texture analysis. The Riesz features are easily extendable to three dimensions [15] to reduce the confusions between micronodules and healthy bronchovascular structures in isotropic multidetector CT. The balanced performance achieved among the various classes of lung tissue suggest that the proposed features are efficient to analyze lung tissue patterns for a large variety of ILD diagnoses. The ability of the Riesz features to adapt to lung textures is illustrated in Fig. 4 where templates according to the dominant principal components of the feature space are shown. In future work, feature selection and learning methods will be incorporated to promote the most relevant Riesz components and reduce the influence of noise. Thanks to the affine-covariant properties of the proposed methods, they are expected to provide tools for analyzing textures with no prior assumptions on translation, scale and orientation parameters in various applications.

Acknowledgments. This work was supported by the Swiss National Science Foundation (under grants 205321-130046 and PP00P2-123438), the CIBM, and the EU 7th Framework Program in the context of Khresmoi (FP7-257528).

References

1. Tourassi, G.D.: Journey toward computer-aided diagnosis: Role of image texture analysis. *Radiology* 213(2), 317–320 (1999)
2. Berlin, L.: Malpractice issues in radiology. perceptual errors. *AJR Am. J. Roentgenol.* 167(3), 587–590 (1996)
3. Castellano, G., Bonilha, L., Li, L.M., Cendes, F.: Texture analysis of medical images. *Clin. Radiol.* 59(12), 1061–1069 (2004)
4. Webb, W.R., Müller, N.L., Naidich, D.P.: *High-Resolution CT of the Lung*. Lipincott Williams & Wilkins, Philadelphia (2001)
5. Delorme, S., Keller-Reichenbecher, M.A., Zuna, I., Schlegel, W., Van Kaick, G.: Usual interstitial pneumonia: Quantitative assessment of high-resolution computed tomography findings by computer-assisted texture-based image analysis. *Invest. Radiol.* 32(9), 566–574 (1997)
6. Uppaluri, R., Hoffman, E.A., Sonka, M., Hunninghake, G.W., McLennan, G.: Interstitial lung disease: A quantitative study using the adaptive multiple feature method. *Am. J. Respir. Crit. Care Med.* 159(2), 519–525 (1999)
7. Sluimer, I.C., van Waes, P.F., Viergever, M.A., van Ginneken, B.: Computer-aided diagnosis in high resolution CT of the lungs. *Med. Phys.* 30(12), 3081–3090 (2003)
8. Sørensen, L., Shaker, S.B., de Bruijne, M.: Quantitative analysis of pulmonary emphysema using local binary patterns. *IEEE Trans. Med. Imaging* 29(2), 559–569 (2010)
9. Depeursinge, A.: Affine-invariant texture analysis and retrieval of 3D medical images with clinical context integration. PhD thesis, University of Geneva (2010)
10. Xu, Y., Sonka, M., McLennan, G., Guo, J., Hoffman, E.A.: MDCT-based 3D texture classification of emphysema and early smoking related lung pathologies. *IEEE Trans. Med. Imaging* 25(4), 464–475 (2006)
11. Caban, J.J., Yao, J., Avila, N.A., Fontana, J.R., Manganiello, V.C.: Texture-based computer-aided diagnosis system for lung fibrosis. In: Giger, M.L., Karssemeijer, N. (eds.) *Medical Imaging 2007: Computer-Aided Diagnosis*, vol. 6514. SPIE, San Jose (2007)
12. Haralick, R.M., Shanmugam, K., Dinstein, I.: Textural features for image classification. *IEEE Trans. Syst. Man and Cybern.* 3(6), 610–621 (1973)
13. Depeursinge, A., Sage, D., Hidki, A., Platon, A., Poletti, P.A., Unser, M., Müller, H.: Lung tissue classification using Wavelet frames. In: *Conf. Proc. IEEE Eng. Med. Biol. Soc.*, pp. 6259–6262. IEEE Computer Society, Los Alamitos (2007)
14. Sørensen, L., Lo, P., Ashraf, H., Sporning, J., Nielsen, M., de Bruijne, M.: Learning COPD sensitive filters in pulmonary CT. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) *MICCAI 2009*. LNCS, vol. 5762, pp. 699–706. Springer, Heidelberg (2009)
15. Unser, M., Van De Ville, D.: Wavelet steerability and the higher-order Riesz transform. *IEEE Trans. Imag. Proc.* 19(3), 636–652 (2010)
16. Depeursinge, A., Vargas, A., Platon, A., Geissbuhler, A., Poletti, P.A., Müller, H.: Building a reference multimedia database for interstitial lung diseases. *Comput. Med. Imaging. Graph.* (to appear)
17. Stein, E.M., Weiss, G.: *Introduction to Fourier analysis on Euclidean spaces*. Princeton University Press, Princeton (1971)

Fast Multiple Organ Detection and Localization in Whole-Body MR Dixon Sequences

Olivier Pauly¹, Ben Glocker², Antonio Criminisi², Diana Mateus¹,
Axel Martinez Möller³, Stephan Nekolla³, and Nassir Navab¹

¹ Computer Aided Medical Procedures, Technische Universität München, Germany
{pauly, mateus, navab}@cs.tum.edu

² Microsoft Research Ltd., CB3 0FB, Cambridge, UK
{glocker, antcrim}@microsoft.com

³ Nuklearmedizin, Klinikum Rechts der Isar, Technische Universität München, Germany
{a.martinez-moller, s.nekolla}@lrz.tu-muenchen.de

Abstract. Automatic localization of multiple anatomical structures in medical images provides important semantic information with potential benefits to diverse clinical applications. Aiming at organ-specific attenuation correction in PET/MR imaging, we propose an efficient approach for estimating location and size of multiple anatomical structures in MR scans. Our contribution is three-fold: (1) we apply supervised regression techniques to the problem of anatomy detection and localization in whole-body MR, (2) we adapt random ferns to produce multi-dimensional regression output and compare them with random regression forests, and (3) introduce the use of 3D LBP descriptors in multi-channel MR Dixon sequences. The localization accuracy achieved with both fern- and forest-based approaches is evaluated by direct comparison with state of the art atlas-based registration, on ground-truth data from 33 patients. Our results demonstrate improved anatomy localization accuracy with higher efficiency and robustness.

1 Introduction

Following the success of combined PET/CT, the possibility of combining PET with MRI has gained increased interest, as significant advantages are expected compared to PET/CT for many imaging tasks in neurology, oncology and cardiology [1]. However, before its introduction in the clinical practice, a technical challenge impacting the quality of PET/MR imaging needs to be solved: the attenuation correction of 511 keV photons according to the radiodensity of the tissues. While in PET/CT [2], radiodensity information provided by CT at X-ray energies can be converted into attenuation information, MR does not provide any information on the tissue density. Therefore, methods have been investigated to generate an attenuation correction map directly from MR. For brain imaging, atlas-based solutions using registration were evaluated in [3,4]. For whole-body imaging, different approaches based on the classification of tissues into 4 classes (background, lungs, fat, and soft tissue) have been investigated, for instance in [5]. While previous methods showed promising results for attenuation correction of whole body imaging with PET/MR, they propose only a coarse tissue classification, not accounting for organ-specific attenuation and for the attenuation introduced by bones.

To further improve the quality of whole-body PET data reconstruction, we aim at generating organ-specific attenuation information directly from MR. Therefore, the position of the organs which impact the attenuation of photons need to be known. This paper presents an approach for simultaneously localizing multiple organs in multi-channel whole-body MR. It builds upon state-of-the-art non-linear regression techniques. We adapt random ferns for regression and compare them to random regression forests. Experiments on 33 patient scans demonstrate better performance than atlas-based techniques in terms of accuracy, speed, and robustness.

2 Related Work

Classical object detection algorithms are based on sliding windows and classifiers whose role is to predict whether a voxel belongs to the object of interest or not. In [6], Viola and Jones introduced a fast detection approach based on a cascade of classifiers trained using Adaboost. Built as a succession of classifiers taking sequentially more and more features into account. This approach achieved impressive performance for real-time face detection. In medical applications, there has been an increasing interest in regression-based solutions for organ localization. Since the human body consists of a specific arrangement of organs and tissues, it can be expected that voxels, based on their contextual information, can predict the surrounding anatomy. For instance, if the neighborhood of a voxel shows an appearance which is typical of heart tissue, besides the position of the heart, this voxel can provide an estimate of position of the nearby lungs. In [7], Zhou *et al.* introduced an approach based on boosting ridge regression to detect and localize the left ventricle (LV) in cardiac ultrasound 2D images. There, the learned function predicts the relative position, scale and orientation of the LV based on Haar-like features computed on 2D images. Impressive results are demonstrated on echocardiogram sequences. To detect and localize the heart chambers in 3D cardiac CT, Zheng *et al.* proposed in [8] an approach called marginal space learning (MSL). To break down the complexity of learning directly in the full 3D similarity transformation space, the authors demonstrate that training a classifier on projections of the original space effectively reduces the search space. Using this idea, they build a cascade of classifiers based on probabilistic boosting tree (PBT) to predict first the position, then the position-orientation and finally the full 3D pose. In [9], the authors push this idea further to non-rigid marginal space learning using statistical shape models. Although these approaches have shown very good performance on CT scans, building such a cascade of classifiers is a computationally intensive learning procedure which requires large training sets. In this paper we avoid intensive training by building a single regressor predicting simultaneously the position of multiple organs. In [10], Criminisi *et al.* proposed a regression approach based on random forests for the localization of organs in 3D CT scans. The authors showed that their method achieves better performance than atlas registration, and this, while benefiting from fast training and testing. While in [10], the authors could rely on absolute radiodensity values provided by CT, here, we deal with MR images which provide only relative values and suffer from field inhomogeneities. To tackle this challenging problem, we adapt the regression forest framework by introducing 3D LBP descriptors. Additionally, we implement a random ferns approach [11]

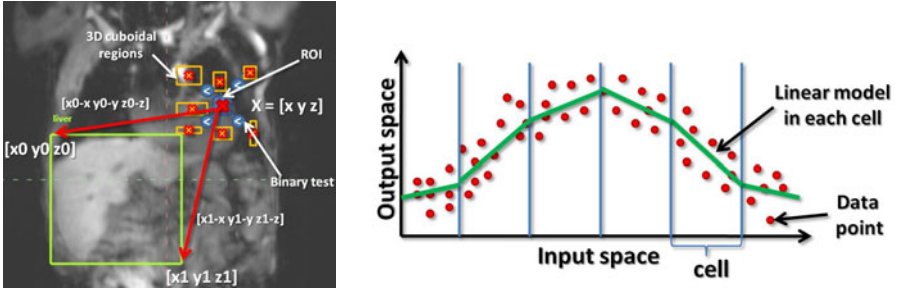


Fig. 1. (Left) Each voxel predicts its relative displacement to all organ bounding boxes represented here as a green box. Multi-scale textural information is extracted using LBP-like feature representation computed over 3D cuboidal regions. (Right) 1D regression example: data samples in red, blue lines represent the partition built over the input feature space. In each cell, simple linear models (in green) are fitted to the points. Their combination over the full space results in a complex non-linear predictor.

and compare it with forests. Both regression techniques are evaluated and compared to an atlas-based registration approach.

3 Proposed Method

This section describes details of our organ detection and localization approach. First, we cast this problem as a regression task. Second, we introduce our feature representation based on water and fat channels computed from MR Dixon sequences. Third, we present regression forests and explain how to adapt random ferns for regression. Finally, we show how to combine voxel predictions to localize all organs of interest in one shot.

3.1 Problem Statement

In our framework each voxel votes for the *relative position* of all organs of interest. The individual votes will produce very noisy predictions. But the probabilistically weighted combination of all votes will produce an accurate output (see Fig. 1). Since the relative displacements we want to predict are continuous values, we use a *regression* paradigm. In the following, we introduce new features based on the fat and water MR Dixon channels, and then present a new non-linear regression approach based on random ferns. Next, we introduce the general problem of organ localization as a regression task.

Input Space: Formally, let us consider the water W and fat F channels computed from MR Dixon sequences defined by the two intensity functions $I^{(W)}, I^{(F)} : \Omega \rightarrow \mathbb{R}$, $\Omega \subset \mathbb{R}^3$ being the image domain. While $\mathbf{x} = [x, y, z]$ represents a voxel location in this domain Ω , $\Psi(I^{(W)}, I^{(F)}, \mathbf{x}) = \mathbf{X}$ denotes a function mapping the voxel location to a feature space according to both intensity functions $I^{(W)}$ and $I^{(F)}$. The role of feature representation \mathbf{X} is to encode contextual information in the neighborhood of location \mathbf{x} computed using $I^{(W)}$ and $I^{(F)}$. \mathbf{X} is the input of our regression function.

Output Space: Let us now consider a set of K organs of interest contained in bounding boxes $\mathcal{O} = \{O_1, \dots, O_k, \dots, O_K\}$. Each bounding box O_k is represented by a vector $O_k = [x_k^0, y_k^0, z_k^0, x_k^1, y_k^1, z_k^1]$. The relative displacement \mathbf{v}_k between voxel location \mathbf{x} and bounding box O_k is parametrized as:

$$\mathbf{v}_k = [x_k^0 - x, y_k^0 - y, z_k^0 - z, x_k^1 - x, y_k^1 - y, z_k^1 - z] \quad (1)$$

We denote by $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_k, \dots, \mathbf{v}_K]$ the vector containing relative displacements between \mathbf{x} and *all* organs of interest. \mathbf{V} is the output of our regression function. On Fig. 1, these relative displacements between the point \mathbf{x} and the liver bounding box are represented by the red arrows. Here, we consider the following organs: head, left lung, right lung, heart and liver.

Regression: We assume a training set $(\mathbf{X}^{(n)}, \mathbf{V}^{(n)})_{n=1}^N$ computed over a set of N patient MR volumes. We could think of modeling the posterior distribution $p(\mathbf{V}|\mathbf{X})$ linking the input and output spaces. However, in such high-dimensional feature spaces, modeling the posterior distribution directly is very difficult. To break down the complexity of this problem, we can first subdivide the input feature space by building a partition \mathcal{P} over it. Indeed, by subdividing the feature space, we obtain cells containing data points which are easier to model even with simple mathematical models such linear or constant functions. As illustrated by the low-dimensional toy example on Fig. 2 (right), the combination of these models over the whole partition results then in a complex non-linear model. Formally, \mathcal{P} is defined as an ensemble of T cells $\mathcal{P} = \{\mathcal{C}_t\}_{t=1}^T$. With \mathcal{P} given, we propose to model the posterior in each cell \mathcal{C}_t as follows:

$$p(\mathbf{V}|\mathbf{X} \in \mathcal{C}_t, \mathcal{P}) = \mathcal{N}_t(\mathbf{X}|\mu_t, \Sigma_t) \quad (2)$$

where \mathcal{N}_t is a multivariate Gaussian distribution whose parameters are estimated during the training phase. In fact, this choice permits to model the full distribution as a piecewise Gaussian distribution. In contrast to fitting a Gaussian mixture model, partitioning is here performed in the input feature space and not in the output space. Moreover, in the case of trees, this partitioning is performed hierarchically. Based on this, we can model the probability distribution of \mathbf{V} over the full feature space according to partition \mathcal{P} as:

$$p(\mathbf{V}|\mathcal{P}) = \sum_{t=1}^T p(\mathbf{V}|\mathcal{C}_t, \mathcal{P})p(\mathcal{C}_t) \quad (3)$$

Clearly, the quality of the posterior approximation depends on the partition \mathcal{P} . If its number of cells T is low, then the posterior approximation will be very rough. On the other hand, if T is high, each cell will include few training points. In this case, the partition \mathcal{P} tends to overfit the training data and suffers from poor generalization. In [12], Breiman demonstrates that replacing a single partition \mathcal{P} with an ensemble of independent random partitions $\{\mathcal{P}_z\}_{z=1}^Z$ leads to an ensemble regressor achieving better generalization. In this paper, we apply regression forests and adapt random ferns for regression to construct multiple independent partitions $\{\mathcal{P}_z\}_{z=1}^Z$. The posterior estimates from the different partitions of the ensemble are then combined using averaging. Finally, we can estimate in one shot the position of all organs of interest contained in

variable $\hat{\mathbf{V}}$ using the mathematical expectation: $\hat{\mathbf{V}} = \int_{\mathbf{V}} \mathbf{V} p(\mathbf{V}) d\mathbf{V}$. Before going into the details of our regression techniques in section 3.3, let us first describe the feature representation we use in the problem of organ localization in MR Dixon sequences.

3.2 Feature Representation

As described in [13], MR Dixon imaging techniques are based on the one shot acquisition of a so-called “in phase” scan where water and fat signals are in-phase and an “opposite phase” scan where water and fat signals are 180° out-of-phase. Using these 2 scans from the same patient, water and fat signals can be separated to construct a water $I^{(W)}$ and a fat $I^{(F)}$ channel. Since these 2 channels are perfectly registered, we propose to take advantage from their complementary nature and design a feature representation based on both water and fat information. While in CT intensity information is directly related to the underlying tissue distribution, MR intensity information is not absolute and suffers from variability between different images. For this reason, we will not rely on intensities as in [10], but on textural information by employing Local Binary Patterns (LBP) [14]: we propose to extract textural context variations at different scales (see Fig. 1). Let us consider a 3D region $\mathcal{R}_{\mathbf{x}}^s$ at scale s centered on voxel location \mathbf{x} and a set $\{\mathbf{N}_{\mathbf{x}}^{s,q}\}_{q=1}^Q$ of Q 3D asymmetric cuboidal regions having different sizes, orientations and offsets in the neighborhood of \mathbf{x} . Using this, we can extract two binary feature vectors $\mathbf{X}_s^{(W)}$ and $\mathbf{X}_s^{(F)}$ from the two channels where each entry is the result of the following binary test comparing average intensities within regions $\mathbf{N}_{\mathbf{x}}^{s,q}$ and $\mathcal{R}_{\mathbf{x}}^s$:

$$\mathbf{X}_s^{(i)}[q] = \frac{1}{|\mathbf{N}_{\mathbf{x}}^{s,q}|} \sum_{\mathbf{x}' \in \mathbf{N}_{\mathbf{x}}^{s,q}} I^{(i)}(\mathbf{x}') < \frac{1}{|\mathcal{R}_{\mathbf{x}}^s|} \sum_{\mathbf{x}' \in \mathcal{R}_{\mathbf{x}}^s} I^{(i)}(\mathbf{x}'), \quad (4)$$

and this, $\forall q \in \{1, \dots, Q\}$ and $i \in \{W, F\}$. Repeating this operation at several scales results in two feature vectors $\mathbf{X}^{(W)}$ and $\mathbf{X}^{(F)}$ describing the multi-scale textural context for both channels in the neighborhood of voxel location \mathbf{x} . Since $\mathbf{X}^{(W)}$ and $\mathbf{X}^{(F)}$ are binary vectors, they can be further encoded to reduce their dimensionality. Finally, they are concatenated in one feature vector: $\mathbf{X} = [\mathbf{X}^{(W)}, \mathbf{X}^{(F)}]$.

3.3 Ensemble Regression Approaches

This section explains how to use forests and ferns to efficiently partition the input data. While regression forests have been used for detecting organs in CT [10], there exists little work on ferns-based regression. In [15], Dollar *et al.* use a ferns-based regressor in a cascade fashion for pose detection of objects in 2D images. In contrast, we use a single ensemble regressor.

Forests and ferns: *Random forests* [12] are constructed as an ensemble of independent random trees. Each tree is a set of decision functions that split feature vectors at each node towards the left or the right branch. *Random ferns*, later introduced by Özüysal *et al.* [11], are an ensemble of *constrained* trees. While a tree applies a different decision function at each node, a fern systematically applies the same decision function for each node of the same level. Results of these random tests are finally stored as binary values.

Ferns benefit from a more compact and simple structure for an accuracy that is similar to those of random trees [11]. In fact, while $2^{(N-1)}$ operations are needed to grow a tree of 2^N leaves, only N operations are needed to grow a fern of 2^N leaves.

Let us now describe these partitioning approaches more formally. We denote by $\mathcal{F} = \{\mathbf{F}^{(z)}\}_{z=1}^Z$ an ensemble of trees or ferns. Each element of this ensemble induces an independent partition $\mathcal{P}^{(z)} = \{\mathcal{C}_1^{(z)}, \dots, \mathcal{C}_T^{(z)}\}$ of the input feature space. Each tree or fern $\mathbf{F}^{(z)}$ is defined as a set of L nodes, each node being equipped with a linear function f_l and an associated threshold τ_l , $l \in \{1, \dots, L\}$. The output of evaluating a pair $(f_l^{(z)}, \tau_l^{(z)})$ on a visual feature vector \mathbf{X} is binary, that is $f_l^{(z)}(\mathbf{X}, \tau_l^{(z)}) : \mathbf{X} \mapsto \{0, 1\}$. While in a tree, the output value denoted by $b_l^{(z)}$ decides whether \mathbf{X} gets pushed towards the left or the right branch, in a fern, \mathbf{X} is evaluated at all nodes. The corresponding outputs are then stored in binary vector $\mathbf{b}^{(z)} = [b_1^{(z)}, \dots, b_l^{(z)}, \dots, b_L^{(z)}]^\top$. In the end, while \mathbf{X} is pushed through the whole tree until it reaches a leaf (which is a cell $\mathcal{C}_t^{(z)}$ of the partition $\mathcal{P}^{(z)}$), in a fern, the full vector $\mathbf{b}^{(z)}$ encodes the cell index of the partition where the vector falls.

Training/Testing: During the training of a fern, the whole training data is used at each node. This is in contrast to trees where only a subset is considered at each node. If we consider a training set $(\mathbf{X}^{(n)}, \mathbf{V}^{(n)})_{n=1}^N$ computed over a set of different patient scans, all feature vectors are pushed through the ferns ensemble and fall into the cells of the different partitions. Finally, the parameters of each Gaussian can be estimated for each cell $\mathcal{C}_t^{(z)}$ using the subset $\{\mathbf{V}^{(n)} | \mathbf{X}^{(n)} \in \mathcal{C}_t^{(z)}\}_{n=1}^N$ of training data that fell into $\mathcal{C}_t^{(z)}$. In the current paper, we do not use optimization in the construction of our ferns regressor, *i.e.* the linear functions and their associated thresholds are chosen randomly. While this permits to have a very fast training procedure, it provides independency from the training set. This can be an advantage for instance in the case of noisy data. Once the training has been performed, all node functions and thresholds are frozen. During the test phase, an unseen data point \mathbf{X} is pushed through the whole ensemble until it reaches a cell in each partition. Then, each cell contributes to the final prediction using its stored Gaussian model as seen in section 3.1. Next, we describe how to combine the predictions to localize all organs of interest.

3.4 Anatomy Localization

Let us consider the water $I^{(W)}$ and fat $I^{(F)}$ channels of an unseen patient. From both channels, a set of feature vectors $\{\mathbf{X}^{(n)}\}_{n=1}^N$ is extracted from voxel locations $\{\mathbf{x}^{(n)}\}_{n=1}^N$. By pushing this set of feature vectors through the regression ensemble, predictions $\{\hat{\mathbf{V}}^{(n)}\}_{n=1}^N$ are computed as described in section 3.1. They correspond to the relative displacements $\hat{\mathbf{V}}^{(n)} = [\hat{v}_1, \dots, \hat{v}_k, \dots, \hat{v}_K]$ between each location $\mathbf{x}^{(n)} = [x^{(n)}, y^{(n)}, z^{(n)}]$ and all organ bounding boxes $\mathcal{O} = \{O_1, \dots, O_k, \dots, O_K\}$. The bounding box of organ O_k can be finally estimated as follows:

$$O_k = \sum_{n=1}^N w_n \left(\hat{\mathbf{v}}_k^{(n)} + \left[\mathbf{x}^{(n)}, \mathbf{x}^{(n)} \right] \right) \quad (5)$$

where each w_n weights the contribution of voxels according to the confidence of their predictions. Note that $\sum_{n=1}^N w_n = 1$. In this paper, we discard contributions having low confidence and perform averaging on the remaining predictions.

4 Experiments and Results

In this section, we compare our approaches based on regression forests and random ferns with the current state-of-the-art multi-atlas registration.

Data: Our dataset currently consists of scans from 33 patients who underwent a 3-Tesla whole-body MR Dixon sequence. All patients have cancer (mostly neck, lung, liver cancer) and show a high variability in their anatomy partially due to their disease. For the detection and localization of organs, we use the water and fat channels. In each scan, we manually delineated the bounding boxes for following organs: head, left lung, right lung, liver and heart. The size of the volumes are $192 \times 124 \times 443$ and the pixel spacing is $2.6 \times 2.6 \times 2.6$ mm.

Regression approach: 100 runs of cross-validation experiments have been conducted where each experiment consists of a training phase on 20 patients chosen randomly and a test phase on the 13 remaining patients. For both forests and ferns, all parameters (number of trees/ferns and tree depth/number of nodes) have been tuned by performing grid-search within the same range for both techniques. Note that node optimization has been performed for random forests based on information gain (cf. [10]). For prediction, each fourth pixel is used and described using 3D LBPs computed over 26 cuboidal regions chosen at 3 different scales.

Multi-atlas registration: 100 runs of cross-validation experiments have been performed. Each experiment is defined as follows: a set of 20 patients are chosen randomly as multi-atlas database and 1 patient is randomly chosen as test case. All 20 patients from the database are registered to the test patient using affine registration. Then, using the ground truth position of the bounding boxes of the test patient (which is not available in reality), we evaluate the theoretical lower and upper bounds of the error by using the patients in the database who provide the lowest and highest localization error. The mean error is computed over the whole database.

Results: Results reported on Tab.1 shows that we achieve an accuracy which is better than the “best case” atlas accuracy, while providing an increased robustness. Taking a look at the localization error per organ, one can notice that the lowest error for our approach is achieved for the localization of the head, which is due to the fact that the head is surrounded by a lot of air which makes it easier to localize. While the heart shows second lowest error, lungs and liver were more difficult to localize. This is mainly due to the high inter-patient variability of the shape of these organs. The best results were obtained with 14 ferns/6 nodes for random ferns, and 6 trees/depth of 8 for regression forests. On a laptop with MATLAB 64 Core Duo 2.4 GHz, the training/testing time on

Table 1. Comparative results: Compared to atlas-based method, our approaches based on random ferns and forests achieve better accuracy and lower uncertainty

MEAN LOCALIZATION ERRORS (mm)						
Organs	Head	Left lung	Right lung	Liver	Heart	Overall
Random ferns	9.82 ± 8.07	14.95 ± 11.35	16.12 ± 11.73	18.69 ± 13.77	15.17 ± 11.70	14.95 ± 11.33
Random forests	10.02 ± 8.15	14.78 ± 11.72	16.20 ± 12.14	18.99 ± 13.88	15.28 ± 11.89	15.06 ± 11.55
Atlas lower bound	18.00 ± 14.45	14.94 ± 11.54	15.02 ± 13.69	18.13 ± 16.26	13.31 ± 11.03	15.88 ± 13.40
Atlas upper bound	70.25 ± 34.23	60.78 ± 29.47	63.95 ± 30.13	70.59 ± 32.88	60.38 ± 28.90	65.19 ± 31.12
Atlas Mean	35.10 ± 13.17	30.41 ± 11.39	29.85 ± 12.62	31.74 ± 13.49	29.82 ± 12.23	31.38 ± 12.58

20/13 patients is 0.7/0.5 s for random ferns. Random Forests need 25/1 s. Concerning atlas registration, each single affine registration needs 12.5 s. To conclude, our approach provides a fast and robust solution for organ detection and localization and thus fulfills our requirements towards organ-specific attenuation map.

5 Conclusion

Our contribution is a supervised regression approach based on random ferns and random forests to detect and localize in one shot multiple organs in whole-body multi-channel MR images. Experiments conducted on a dataset of 33 patients show that our approach achieves an accuracy which is better than atlas-based methods, while providing higher robustness (lower uncertainty) and faster training/prediction times. Furthermore, this approach can be also useful to integrate semantic information *i.e.* incorporating organ labels in further applications such as registration, image navigation or image retrieval. In future work, we plan to investigate the online performance of the proposed approach to enable a fast updating of our organ localization system, and then we will move towards the construction of organ-specific attenuation correction maps.

References

1. Judenhofer, M.S., et al.: Simultaneous pet-mri: a new approach for functional and morphological imaging. *Nature Medicine*, 459–465 (2008)
2. Kinahan, P., Hasegawa, B., Beyer, T.: X-ray-based attenuation correction for positron emission tomography/computed tomography scanners. *Semin. Nuc. Med.* (2003)
3. Kops, E.R., Herzog, H.: Template-based attenuation correction of PET in hybrid MR-PET. *Journal of Nuclear Medicine* (2008)
4. Hofmann, M., Steinke, F., Scheel, V., Charpiat, G., Farquhar, J., Aschoff, P., Brady, M., Schölkopf, B., Pichler, B.J.: MRI-Based Attenuation Correction for PET/MRI: A Novel Approach Combining Pattern Recognition and Atlas Registration. *Journal of Nuclear Medicine* (2008)
5. Martinez-Möller, A., Souvatzoglou, M., Delso, G., Bundschuh, R.A., Chefd'hotel, C., Ziegler, S.I., Navab, N., Schwaiger, M., Nekolla, S.G.: Tissue Classification as a Potential Approach for Attenuation Correction in Whole-Body PET/MRI: Evaluation with PET/CT Data. *Journal of Nuclear Medicine* 50, 520–526 (2009)
6. Viola, P., Jones, M.J.: Robust real-time face detection. *International Journal on Computer Vision* 57, 137–154 (2004)

7. Zhou, S.K., Zhou, J., Comaniciu, D.: A boosting regression approach to medical anatomy detection. In: IEEE Conference on Computer Vision and Pattern Recognition (2007)
8. Zheng, Y., Barbu, A., Georgescu, B., Scheuering, M., Comaniciu, D.: Four-chamber heart modeling and automatic segmentation for 3-D cardiac CT volumes using marginal space learning and steerable features. *IEEE Transactions on Medical Imaging* 27, 1668–1681 (2008)
9. Zheng, Y., Bogdan, C.D.: Marginal Space Learning for Efficient Detection of 2D/3D Anatomical Structures in Medical Images. In: Prince, J.L., Pham, D.L., Myers, K.J. (eds.) *IPMI 2009. LNCS*, vol. 5636, pp. 411–422. Springer, Heidelberg (2009)
10. Criminisi, A., Shotton, J., Robertson, D., Konukoglu, E.: Regression Forests for Efficient Anatomy Detection and Localization in CT Studies. In: Jiang, T., Navab, N., Pluim, J., Viergever, M. (eds.) *MICCAI 2010 Workshop in Medical Computer Vision* (2010)
11. Özuysal, M., Calonder, M., Lepetit, V., Fua, P.: Fast Keypoint Recognition using Random Ferns. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32, 448–461 (2010)
12. Breiman, L.: Random forests. *Machine Learning* 45, 5–32 (2001)
13. Ma, J.: Dixon Techniques for Water and Fat Imaging. *J. Mag. Res. Im.* (2008)
14. Ojala, T., Pietikäinen, M., Harwood, D.: A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition* 29, 51–59 (1996)
15. Dollar, P., Welinder, P., Perona, P.: Cascaded pose regression. In: IEEE Conference on Computer Vision and Pattern Recognition (2010)

ManiSMC: A New Method Using Manifold Modeling and Sequential Monte Carlo Sampler for Boosting Navigated Bronchoscopy

Xiongbiao Luo¹, Takayuki Kitasaka², 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. This paper presents a new bronchoscope motion tracking method that utilizes manifold modeling and sequential Monte Carlo (SMC) sampler to boost navigated bronchoscopy. Our strategy to estimate the bronchoscope motions comprises two main stages:(1) bronchoscopic scene identification and (2) SMC sampling. We extend a spatial local and global regressive mapping (LGRM) method to Spatial-LGRM to learn bronchoscopic video sequences and construct their manifolds. By these manifolds, we can classify bronchoscopic scenes to bronchial branches where a bronchoscope is located. Next, we employ a SMC sampler based on a selective image similarity measure to integrate estimates of stage (1) to refine positions and orientations of a bronchoscope. Our proposed method was validated on patient datasets. Experimental results demonstrate the effectiveness and robustness of our method for bronchoscopic navigation without an additional position sensor.

1 Introduction

During bronchoscopic interventions, physicians must know the position and orientations of a bronchoscope inside the airway trees, since they usually perform transbronchial lung biopsy (TBLB) to obtain samples of suspicious tumors for the assessment of bronchus and lung cancer. To localize and track the bronchoscope, current state of the art in navigated bronchoscopy includes two main approaches (or a combination of both): (1) image-based algorithms and (2) electromagnetic tracking (EMT). Although these methods proved good performance [1,2,3], it remains challenging to correctly localize the bronchoscope to places where it is exactly observing. Image-based schemes cannot tackle situations where problematic bronchoscopic video images (e.g., bubbles and motion blurring) happen. EMT-based methods often locate bronchoscopes incorrectly under any airway deformation, and accuracy of an EMT sensor measurements is heavily worsened by magnetic field distortion. Furthermore, no matter what approaches are used for bronchoscope motion tracking, they hardly adapt themselves to situation changes (e.g., patient coughing or dynamic errors in EMT outputs) over time during bronchoscopic interventions.

Beyond methods mentioned above, our primary idea of bronchoscope motion estimation is that bronchoscopic video sequences basically consist of three categories: (1) uninformative class, (2) inter-bronchus class, and (3) intra-bronchus class. The first class only includes problematic bronchoscopic frames (also refer to ambiguities) due to specular- or inter-reflection, bubbles, motion blurring, or collision with the bronchial walls. The informative category comprises inter- and intra-bronchus classes, which correspond to bronchoscopic frames with or without folds, and bronchoscopic images with bifurcations, respectively. A bronchoscopic image or scene can always be classified to one of three categories.

Based on these ideas, this work first learns bronchoscopic video manifolds to segment a bronchoscopic video into different clusters where each cluster represents one bronchial branch. We here construct bronchoscopic manifolds based on extending a local and global regressive mapping (LGRM) method [4] to Spatial-LGRM. By embedding an input image into these clusters, we can find an optimal bronchial branch that corresponds to its scene where a bronchoscope is observing. Hence, we can roughly obtain the pose of the camera. We then perform SMC sampling based on a selective image similarity measure to integrate the estimates of Spatial-LGRM-based learning to refine the localization of the bronchoscope.

It is worthwhile to highlight several aspects of our approach as follows. First, we propose a new framework of manifold modeling and SMC sampling to determine localizations of a bronchoscope for navigated bronchoscopy beyond image-based methods and EMT systems. Note that our manifold learning-based method for bronchoscopic scene identification provides an almost real-time means to roughly estimate the position and orientation of a bronchoscope. Next, we construct a new manifold modeling called Spatial-LGRM, which combines pose information to characterize the bronchoscope movements. Last, we introduce SMC sampling to incorporate manifold-based estimates and to tackle situations where ambiguities occur in bronchoscopic videos. Additionally, although we focused on bronchoscope motion tracking, we believe that our framework should also be appropriate to navigate other diagnostic endoscopes, e.g., colonoscope.

2 Bronchoscope Motion Tracking: ManiSMC

2.1 Bronchoscopic Scene Identification

(a) Preprocessing. We segment each 3D CT dataset to obtain bronchial tree structure information \mathcal{B} : $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_u, \dots, \mathbf{b}_k; u = 1, 2, \dots, k\}$, where k is the number of bronchial branches, and \mathbf{b}_u describes the centerline of one bronchial branch with its start position \mathbf{s}_u and end position \mathbf{e}_u ; branch direction \mathbf{d}_u can be computed by $\mathbf{d}_u = \mathbf{e}_u - \mathbf{s}_u$. \mathcal{B} is used to generate training data.

For each input RB image I_t at frame t , we first check whether it is an uninformative frame. In the HSB (HSB: Hue, Saturation, Brightness) color space, we compute hue and brightness deviations between current RB image and a virtual image generated by a virtual camera with an estimated pose inside the airway trees. If hue and brightness deviations are bigger than two predefined thresholds, I_t is considered as an uninformative frame. If I_t is an informative image,

we then perform morphology-based bifurcation detection to determine whether I_t is a bifurcation (intra-bronchus) image. We extract hole regions in I_t : if hole regions is more than two and the distances among these regions are constrained in a domain (here need to set minimum and maximum values for this domain), I_t is an image with observed bifurcations.

(b) Training Data Generation. By bronchial centerline information \mathcal{B} , we generate training data by changing positions and orientations of a virtual camera placed in a pre-built 3D anatomical airway model. Let \mathbf{p} denotes the camera position. Three vectors \mathbf{e}_x , \mathbf{e}_y , and \mathbf{e}_z ($\mathbf{e}_x = \mathbf{e}_z \times \mathbf{e}_y$) describe orientations of the virtual camera. We update positional parameters of the virtual camera by:

$$\mathbf{p} = \mathbf{p}_u^v + \frac{r}{3}\mathbf{e}_y, \quad \mathbf{p}_u^v = \mathbf{s}_u + \alpha_u \mathbf{d}_u, \quad (0, \mathbf{e}_y) = \mathbf{q}_\phi(0, \mathbf{e}_y^0) \mathbf{q}_\phi^{-1}, \quad \mathbf{q}_\phi = (\cos \frac{\phi}{2}, \sin \frac{\phi}{2} \mathbf{e}_z^0), \quad (1)$$

where \mathbf{p}_u^v is the v -th chosen point on the centerline of branch \mathbf{b}_u , r is the radius of \mathbf{b}_u , and $\alpha_u \in (0, 1)$ is constant coefficient. Simultaneously, we change orientational information by updating three vectors using quaternion \mathbf{q} :

$$(0, \mathbf{e}_z) = \mathbf{q}_\omega \mathbf{q}_\theta ((0, \mathbf{e}_z^0) \mathbf{q}_\theta^{-1} \mathbf{q}_\omega^{-1}), \quad (0, \mathbf{e}_y) = \mathbf{q}_\varphi \mathbf{q}_\omega \mathbf{q}_\theta ((0, \mathbf{e}_y^0) \mathbf{q}_\theta^{-1} \mathbf{q}_\omega^{-1} \mathbf{q}_\varphi^{-1}), \quad (2)$$

$$\mathbf{q}_\theta = (\cos \frac{\theta}{2}, \sin \frac{\theta}{2} \mathbf{e}_x^0), \quad \mathbf{q}_\omega = (\cos \frac{\omega}{2}, \sin \frac{\omega}{2} \mathbf{e}_z^0), \quad \mathbf{q}_\varphi = (\cos \frac{\varphi}{2}, \sin \frac{\varphi}{2} \mathbf{e}_y^0), \quad (3)$$

where $\mathbf{e}_z^0 = \mathbf{d}_u / \|\mathbf{d}_u\|$, \mathbf{e}_y^0 is a random vector that satisfies: $\mathbf{e}_y^0 \cdot \mathbf{e}_z^0 = 0$, ω and φ are rotational angles around \mathbf{e}_z^0 , θ is a rotational angle around \mathbf{e}_x^0 .

Generally, we generate two categories training data: (1) inter-bronchus and (2) intra-bronchus, by adjusting coefficient α_u : for inter-bronchus (fold) images $\alpha_u \in [0.1, 0.3]$, and intra-bronchus (bifurcation) images with $\alpha_u \in [0.7, 0.9]$.

(c) Learning Bronchoscopic Video Manifolds. After generating training data, we calculate low dimensional embedding spaces or eigenspaces and mapping functions or eigenmaps using LGRM that was proved to provide better performance more than other manifold learning methods in [4].

Suppose n training images $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_i, \dots, \mathbf{x}_n\}$, where $\mathbf{x}_i \in \mathbb{R}^D$, $D = w \times h$, $w \times h$ is the image size. Eigenspace $\mathcal{Y} = \{\mathbf{y}_1, \dots, \mathbf{y}_i, \dots, \mathbf{y}_n\}$ are low dimensional embedding manifolds of \mathcal{X} , where $\mathbf{y}_i \in \mathbb{R}^d$, ($d \ll D$). Manifold learning aims to map \mathcal{X} to \mathcal{Y} by finding eigenmap $\mathcal{M} \in \mathbb{R}^{D \times d}$: $\mathbf{x}_i \rightarrow \mathbf{y}_i, \mathbb{R}^D \rightarrow \mathbb{R}^d$. Note that \mathbf{X} , \mathbf{Y} , and \mathbf{M} represent matrices of \mathcal{X} , \mathcal{Y} , and \mathcal{M} .

In LGRM, since each image \mathbf{x}_i is first transformed to a Hilbert space \mathcal{H} and assumes that there exists a linear projection between \mathcal{H} and \mathbb{R}^d , for any \mathbf{x}_i , its low dimensional embedding \mathbf{y}_i satisfies:

$$\mathbf{y}_i = \psi(\mathbf{M})^T \psi(\mathbf{x}_i) + \mathbf{N}, \quad (4)$$

where $\psi(\mathbf{M})^T$ maps \mathbf{M} from \mathcal{H} to \mathbb{R}^d and \mathbf{N} is a residual term.

Specially, finding \mathbf{Y} and \mathbf{M} can be formulated the following optimization problem in terms of LGRM [4]:

$$\min_{\mathbf{Y}^T \mathbf{Y} = \mathbf{I}} \text{Tr} \left[\mathbf{Y}^T (\mathbf{L}_l + \mu \mathbf{L}_g) \mathbf{Y} \right], \quad (5)$$

where Tr is the trace operator, \mathbf{I} is an identity matrix, $(\mathbf{L}_l + \mu\mathbf{L}_g)$ is the Laplacian matrix that is the key component in manifold learning, and μ is a constant. \mathbf{L}_l that preserves the local manifold structure is calculated by Laplacian Eigenmaps [5]. \mathbf{L}_g denotes a kernelized global regression regularization, which is the main different part from other manifold learning algorithms; it can be computed by $\mathbf{L}_g = \xi\mathbf{H}(\mathbf{H}\mathbf{K}\mathbf{H} + \xi\mathbf{I})^{-1}\mathbf{H}$, where ξ is a regularization parameter, \mathbf{H} is the global centering matrix, and \mathbf{K} is the kernel matrix with its component $K_{ij} = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|/\sigma^2)$ between two training images \mathbf{x}_i and \mathbf{x}_j .

Performing eigen decomposition on $(\mathbf{L}_l + \mu\mathbf{L}_g)$, we can obtain eigenspace \mathbf{Y} . Seeking partial derivatives of Eq. 5, $\psi(\mathbf{M})$ and \mathbf{N} can be determined by 4:

$$\psi(\mathbf{M}) = \psi(\mathbf{X})\mathbf{H}(\mathbf{H}\psi(\mathbf{X})^T\mathbf{X}\mathbf{H} + \xi\mathbf{I})^{-1}\mathbf{Y}, \quad \mathbf{N} = \frac{1}{m}\mathbf{Y}^T\mathbf{1}_m - \frac{1}{m}\mathbf{M}^T\psi(\mathbf{X})\mathbf{1}_m, \quad (6)$$

when constant m reckons on \mathbf{L}_l and $\mathbf{1}_m \in \mathbb{R}^m$ is a vector of one.

After obtaining eigenspace \mathbf{Y} and eigenmap \mathbf{M} , for any input RB image I_t , its embedding \mathbf{y}_t can be easily obtained by Eqs. 4 and 6:

$$\mathbf{y}_t = (\psi(\mathbf{X})\mathbf{H}(\mathbf{H}\psi(\mathbf{X})^T\mathbf{X}\mathbf{H} + \xi\mathbf{I})^{-1}\mathbf{Y})^T\psi(\mathbf{x}_t) + \frac{1}{m}\mathbf{Y}^T\mathbf{1}_m - \frac{1}{m}\mathbf{M}^T\psi(\mathbf{X})\mathbf{1}_m. \quad (7)$$

However, LGMR only use intensity information of training images. This means \mathbf{L}_l and \mathbf{L}_g (or \mathbf{K}) only preserve intensity information of training images. From our experiences, intensity of inter-bronchus (fold) images are quite similar, although they may generate from totally different observation positions and orientations. This results in similar embedding representations of inter-bronchus images in \mathbf{Y} ; it may collapse clusters to wrongly identify bronchoscopic scenes and incorrectly estimate bronchoscope localizations.

To overcome such a drawback of LGRM, we extend it to Spatial-LGRM that integrates spatial information included camera position \mathbf{p} and orientation matrix $\mathbf{r}(\mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z)$, i.e., we add \mathbf{p} and $\mathbf{r}(\mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z)$ to \mathbf{y}_i w.r.t \mathbf{x}_i :

$$(\mathbf{y}_i) \rightarrow \mathbf{y}_i^s = (\mathbf{y}_i, \mathbf{p}^i, \mathbf{e}_x^i, \mathbf{e}_y^i, \mathbf{e}_z^i), \quad \mathbf{Y} \rightarrow \mathbf{Y}^s, \quad \mathbb{R}^d \rightarrow \mathbb{R}^{(d+12)}. \quad (8)$$

Finally, we obtain eigenspace \mathbf{Y}^s and eigenmap \mathbf{M} for RB scene clustering.

(d) Bronchoscopic Scene Clustering. After preprocessing input RB image I_t , we cluster I_t to recognize current bronchoscopic scene by: (a) embedding I_t to \mathbf{Y}^s using \mathbf{M} in terms of Eq. 7 and obtain \mathbf{y}_t^s , (b) calculating Euclidean distance C_{t-1}^i between estimated position \mathbf{p}_{t-1} of I_{t-1} and $\mathbf{p}^i \in \mathbf{y}_i^s \in \mathbf{Y}^s$: $C_{t-1}^i = \sqrt{\|\mathbf{p}_{t-1} - \mathbf{p}^i\|^2}$, and choosing J_p nearest neighbors for \mathbf{y}_t^s by C_{t-1}^i : $\{\mathbf{y}_i^s\}_{i=1}^{J_p}$, (c) computing orientation deviation O_{t-1}^h between estimated rotation matrix \mathbf{r}_{t-1} of I_{t-1} and $\mathbf{r}^h(\mathbf{e}_x^h, \mathbf{e}_y^h, \mathbf{e}_z^h) \in \mathbf{y}_h^s \in \{\mathbf{y}_i^s\}_{i=1}^{J_p}$: $O_{t-1}^h = \arccos((\text{Tr}(\mathbf{r}^h\mathbf{r}_{t-1}^T) - 1)/2)$, and selecting J_o nearest neighbors by O_{t-1}^h from $\{\mathbf{y}_i^s\}_{i=1}^{J_p}$: $\{\mathbf{y}_h^s\}_{h=1}^{J_o}$, and (d) calculating distance E_t^h between \mathbf{y}_t^s and $\mathbf{y}_h^s \in \{\mathbf{y}_i^s\}_{i=1}^{J_o}$: $E_t^h = \sqrt{\|\mathbf{y}_t^s - \mathbf{y}_h^s\|^2}$. Finally, the output of the cluster is the pose parameters $(\mathbf{p}_*^s, \mathbf{r}_*^s)$ that correspond to the optimal embedding \mathbf{y}_*^s ($\mathbf{y}_*^s \in \{\mathbf{y}_h^s\}_{h=1}^{J_o}$) that is the closest to \mathbf{y}_t^s .

2.2 SMC-Based Bronchoscope Motion Estimation

This section fuses pose parameters $(\mathbf{p}_*^s, \mathbf{r}_*^s)$ estimated from the step of bronchoscopic scene identification to determine current RB pose using a SMC sampler. Since this stage is quite similar to our previous work [6], we here briefly review the processing of SMC-based bronchoscope motion estimation.

Let $\mathbf{Q}_t(\mathbf{p}_t, \mathbf{r}_t)$ with translation \mathbf{p}_t and rotation matrix \mathbf{r}_t denotes transformation matrix from bronchoscope coordinates to CT coordinates at frame t .

We generate set of random samples $\mathcal{S}_t^g = \{(\mathbf{Q}_t^g, w_t^g) : t = 1, 2, \dots, N; g = 1, 2, \dots, M\}$ (N and M are the number of frames and samples, and w_t^g is a sample weight) to approach the posterior distribution of bronchoscope motion \mathbf{Q}_t . These samples are deterministically drifted and stochastically diffused by $\mathbf{Q}_t^g = \mathbf{A}\mathbf{Q}_{t-1}^g + \mathbf{B}n_t^g$, where matrix \mathbf{A} is calculated from $(\mathbf{p}_*^s, \mathbf{r}_*^s)$ and $\mathbf{B}n_t^k$ is a noise term. Next, using \mathbf{Q}_t^g to generate virtual image I_V , we compute weight w_t^g by a selective image similarity measure [1]: $w_t^g = MoMSE(\mathbf{I}_t, I_V(\mathbf{Q}_t^g))$. Finally, in our case, the output parameters $\hat{\mathbf{Q}}_t$ with position and rotation of the SMC sampler for determining the pose of RB frame \mathbf{I}_t can be determined in terms of w_t^g : $\hat{\mathbf{Q}}_t = \max_{\hat{w}_t^g} \{(\mathbf{Q}_t^g, w_t^g)\}$, i.e., sample $\hat{\mathbf{Q}}_t$ with maximal weight \hat{w}_t^g corresponds to the maximal similarity between the current bronchoscopic image and the virtual frame generated by placing a virtual camera with the estimated pose including translation vector and rotation matrix inside the 3D airway tree anatomical model that was constructed by volume rendering techniques.

3 Experiments

For validation of our proposed method, we applied it to five cases of patient datasets that include bronchoscopic video frames and their corresponding 3-D chest CT images. The acquisition parameters of CT images are 512×512 pixels, 72-361 slices, and 1.0-2.0 mm slice thickness.

In Section 2.1(a), after pre-processing CT data, we obtain bronchial branch structure information \mathcal{B} and a 3D anatomical airway model. We generate training images with 30×30 pixels in gray-scale space by adjusting the following parameters in terms of Eqs. 1-3 in Section 2.1(b): for inter-bronchus images, α_u is set to 0.15, 0.20, and 0.25; for intra-bronchus images, α_u is set to 0.80, 0.85, and 0.90; $\phi = 30^\circ$, $\theta = 15^\circ$, $\omega = 0^\circ, 15^\circ, 30^\circ, \dots, 345^\circ$, and $\varphi = 0^\circ, 30^\circ, 60^\circ, \dots, 330^\circ$. For each branch $\mathbf{b}_u \in \mathcal{B}$, we generates 7488 inter-bronchus and 7488 intra-bronchus frames. During learning bronchoscopic video in Section 2.1, we constructed ten-dimensional embedding manifolds ($d = 10$ and $D = 30 \times 30$). Hence, the dimensions of matrices \mathbf{X} , \mathbf{Y} , and \mathbf{M} are 7488×900 , 7488×10 , and 900×10 . We set $\mu = 10^{-4}$, $\xi = 10^{-5}$, and $\sigma = 100$ in Spatial-LGRM according to [4].

Currently, for each patient case, we use six bronchial branches to create bronchoscopic video manifolds. They are (1) TR: trachea, (2) LM: left main bronchus, (3) RM: right main bronchus, (4) LU: left upper lobe bronchus, (5) RU: right upper lobe bronchus, and (6) RT: right trunchus intermedius. Hence, we totally obtain 12 clusters (inter- and intra-bronchus classes) for one patient and each cluster includes one eigenspace and one eigenmap.

After uninformative and bifurcation detections of input RB image I_t in Section 2.1(a), we convert I_t to a gray-scale image (362×370 or 256×263 pixels) and interpolate it to 30×30 pixels. During scene clustering in Section 2.1(d), the parameters of nearest neighbors are set to: $J_p = 200$ and $J_o = 100$. Additionally, to evaluate the successful rate of uninformative, we generate ground truth by manually inspecting real bronchoscopic images from three observers, one bronchoscopist and two scientists. Moreover, the tracking results of our proposed method are also manually and visually examined by the same three experts.

4 Results and Discussion

Table 1 summarizes the processed results of our methods. Detection rates of uninformative and intra-bronchus images are about 76.3% and 89.7%. The successful or correct scene recognition by only using bronchoscopic scene identification (BSI) described in Section 2.1 is about 3839 frames (59.6 %), which was improved to 4522 frames (70.2 %) using ManiSMC (Section 2). Fig. 1 visually compares the processed results of methods of BSI and ManiSMC. Generally, our experimental results demonstrate the effectiveness of ManiSMC that shows a good performance to understand bronchoscopic videos.

However, our method still fails to correctly estimate the bronchoscope localizations. Several reasons must be clarified as follows. First, sometimes uninformative images are wrongly detected, which results in incorrect embedding in manifolds; e.g., an image with bubbles can never find a correct correspondence in eigenspaces. In the future, we will improve uninformative frame detection by the work of Atasoy et al. [7] or the methods presented in [8,9]. Next, intra-bronchus images are wrongly classified. If an inter-bronchus image is detected to be an intra-bronchus one, it will never obtain a correct embedding in manifold clustering. Fig. 2 (a) (*top*) shows a successful detection of uninformative and bifurcation images. In some cases, it is difficult to detect whether an image is bifurcation, e.g., in Fig. 2 (a) (*bottom*), the RB frame is collided with a bronchial wall. Since detections of uninformative and bifurcation images are important to BSI and ManiSMC, we must improve current detection methods to

Table 1. Quantified processed results of bronchoscope motion tracking by visual inspection that manually checks if a RB image is similar to a virtual one

Patient cases	Number of frames	Moving path	Detection rates		Successful frames	
			Uninformative	Bifurcation	BSI	ManiSMC
1	1436	LM→LU	76.2%	89.7%	59.0%	77.0%
2	1685	TR→LM	74.6%	89.2%	51.4%	63.4%
3	1167	TR→RM→LM	70.8%	88.5%	45.8%	53.7%
4	1053	RM→RU	79.3%	90.3%	71.5%	77.5%
5	1101	TR→RM→RT	80.8%	90.8%	76.1%	82.2%
Total	6442		76.3%	89.7%	59.6%	70.2%

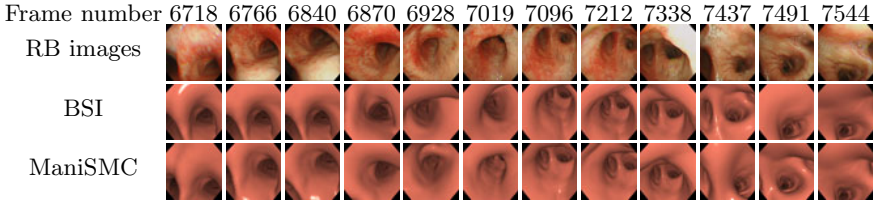


Fig. 1. Visual inspection of processed results of Case 4 by our bronchoscope motion tracking methods. Top row shows selected frame numbers and second row shows their corresponding patient RB images. Other rows display virtual images generated from processed results using methods of BSI and ManiSMC.

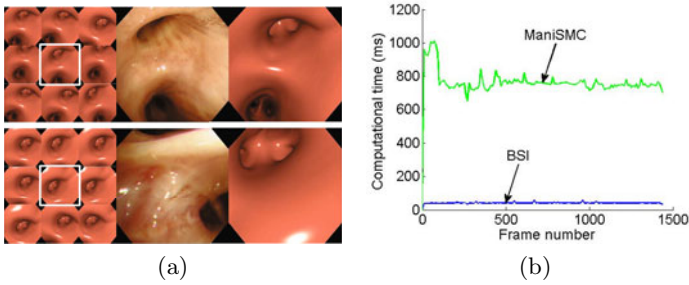


Fig. 2. (a) Examples of bronchoscopic scene identification: successfully recognized (*top*) and unsuccessfully identified (*bottom*) images. They show 9-nearest neighbors (*left*) of BSI, the frame with “□” used to SMC sampling, and RB images (*median*) with corresponding virtual images (*right*) generated by estimates of ManiSMC. (b) Computational times of Case 1. Average processing times of methods of BSI and ManiSMC are about 42 and 750 milliseconds per frame. Note that BSI can process one frame almost in real time (about 25 fps).

further enhance performances of BSI and ManiSMC. Third, similar images such as collision of bronchial walls and convolution of bronchial bifurcations in training data usually confuses clusters to determine accurate embeddings, although correct embeddings are included in the nearest neighbors of input RB images. Forth, we generated training data by updating the virtual camera observation poses in terms of bronchial centerline, i.e., most training images converges the bronchial centerline; however, a bronchoscope is usually not moving along the centerline, which causes actual bronchoscope poses that are difficult to correspond to manifolds. We need to improve the diversity of training data by adding more different virtual camera poses. Moreover, loss of centerline information due to airway segmentation algorithms also contributes to failures of scene identification. Finally, training data were generated from static CT slices that were acquired without airway deformation but bronchoscopic videos include patient breathing or coughing, this also causes unsuccessful bronchoscope tracking. Additionally, computational times of our methods are shown in Fig. 2 (b). Interestingly, BSI can almost process one frame in real time, about 25 frames per second

(fps). ManiSMC needs 0.75 seconds per frame since it requires to compute each sample weight during SMC-based motion estimation that is time-consuming.

5 Conclusion

This work proposed a new method that introduces LGRM-based manifold learning and SMC sampling for bronchoscope motion estimation. We constructed a Spatial-LGRM modeling with camera pose information to learn bronchoscopic video manifolds and use them to identify bronchoscopic video scenes where a bronchoscope is located and observing. Such a method can almost process video frames in real time (about 25 frames per second). By integrating a SMC sampler, our method can tackle situations where ambiguities occur in bronchoscopic videos. We may conclude that our proposed method provides a perspective means to boost bronchoscopic navigation without an additional position sensor.

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. 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)
2. Soper, T.D., Haynor, D.R., Glenn, 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. Luo, X., Feuerstein, M., Sugiura, T., Kitasaka, T., Mori, K., Imaizumi, K., Hasegawa, Y.: Towards hybrid bronchoscope tracking under respiratory motion: evaluation on a dynamic motion phantom. In: Wong, K.H., Miga, M.I. (eds.) *SPIE Medical Imaging 2010*, California, USA, vol. 7625, p. 76251B (2010)
4. Yang, Y., Nie, F., Xiang, S., Zhuang, Y., Wang, W.: Local and global regressive mapping for manifold learning with out-of-sample extrapolation. In: Fox, M., Poole, D. (eds.) *Proceedings of AAAI 2010*, Atlanta, USA, pp. 649–654 (2010)
5. Belkin, M., Niyogi, P.: Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation* 15(5), 1371–1396 (2003)
6. 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)
7. Atasoy, S., Mateus, D., Lallemand, J., Meining, A., Yang, G.Z., Navab, N.: Endoscopic video manifolds. In: Jiang, T., et al. (eds.) *MICCAI 2010*. LNCS, vol. 6362, pp. 437–445. Springer, Heidelberg (2010)
8. Oh, J., Hwang, S., Lee, J., Tavanapong, W., Wong, J., de Groen, P.C.: Informative frame classification for endoscopy video. *MedIA* 11(2), 110–127 (2007)
9. Bashar, M., Kitasaka, T., Suenaga, Y., Mekada, Y., Mori, K.: Automatic detection of informative frames from wireless capsule endoscopy images. *MedIA* 14(3), 449–470 (2010)

Characterizing Pathological Deviations from Normality Using Constrained Manifold-Learning

Nicolas Duchateau, Mathieu De Craene,
Gemma Piella, and Alejandro F. Frangi

Center for Computational Imaging & Simulation Technologies in Biomedicine
(CISTIB) – Universitat Pompeu Fabra and CIBER-BBN, Barcelona, Spain

Abstract. We propose a technique to represent a pathological pattern as a deviation from normality along a manifold structure. Each subject is represented by a map of local motion abnormalities, obtained from a statistical atlas of motion built from a healthy population. The algorithm learns a manifold from a set of patients with varying degrees of the same pathology. The approach extends recent manifold-learning techniques by constraining the manifold to pass by a physiologically meaningful origin representing a normal motion pattern. Individuals are compared to the manifold population through a distance that combines a mapping to the manifold and the path along the manifold to reach its origin. The method is applied in the context of cardiac resynchronization therapy (CRT), focusing on a specific motion pattern of intra-ventricular dyssynchrony called septal flash (SF). We estimate the manifold from 50 CRT candidates with SF and test it on 38 CRT candidates and 21 healthy volunteers. Experiments highlight the need of nonlinear techniques to learn the studied data, and the relevance of the computed distance for comparing individuals to a specific pathological pattern.

1 Introduction

By definition, a disease is an impairment of the normal condition of an organism. Considering different grades of the same disease as progressive deviations from normality addresses therefore the understanding of this disease and facilitates its diagnosis in a given patient. This approach is particularly of interest for cardiac resynchronization therapy (CRT), where the definition of relevant criteria for selecting candidates likely to respond to the therapy is still a topic under active debate [5]. The advantages of considering specific groups of mechanical dyssynchrony in the selection process were recently discussed in [11]. Each of these groups corresponded to one pathological pattern of myocardial motion and deformation with different grades of abnormality with respect to a healthy cardiac function. However, the approach lacks of reproducible tools for the grading of a given pathological pattern within a population and for the quantitative comparison of individuals to each of these specific populations. The aim of this paper is to demonstrate the relevance of describing each pathological pattern as a deviation from normality along a manifold structure, allowing the computation of an appropriate distance between individuals and each pathological pattern.

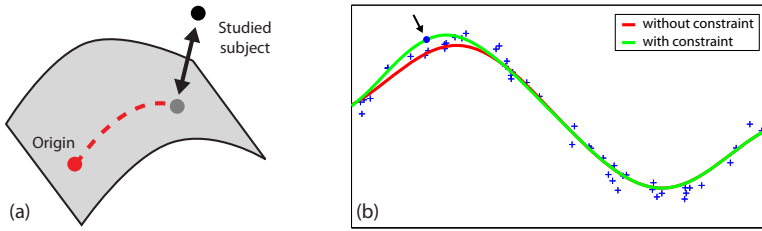


Fig. 1. (a) Distance proposed in this paper, combining a mapping to the manifold and the path along the manifold to reach its origin. (b) Interpolation of a 1D synthetic dataset using inexact matching, before and after the addition of a constraint forcing the curve to pass by the point indicated by the black arrow.

The definition of an optimal space for the comparison of different populations was already addressed by dimensionality reduction techniques, such as principal component analysis (PCA), Kernel-PCA [10], principal geodesic analysis (PGA) [4], linear discriminant analysis [9] or multivariate statistics [14]. Nonetheless, the flexibility of these techniques is limited when a new subject or a new population is added to the existing dataset, as dimensionality reduction is applied to the whole set of studied subjects. In addition, the dimensionality reduction could be biased towards certain populations if they show higher variability. An alternative for moving beyond these limitations consists in separating the analysis for each coherent group of subjects. Using PCA, Kernel-PCA or PGA within a given population is of limited interest for our application, as we target the comparison of individuals to the whole population and not just to its mean or centroid. The comparison of an individual to its k -nearest neighbors (k -NN) does not take into account the local topology of the dataset and assimilates all distances to Euclidean distances [13] [7]. In contrast, manifold learning techniques intrinsically take into account this geometry, and allow relevant comparison of individuals to the studied population through the use of a mapping distance. This mapping results from the “pre-image problem,” used in the literature for denoising [10] [8], segmentation [3], face recognition [15] and regression [1]. A distance based on this mapping mechanism was introduced in [6], but its use was limited to the estimation of reconstruction errors inherent to a reduction of dimensionality.

In this paper, we extend manifold-learning techniques to embed the definition of a relevant origin within the manifold. We propose a distance for comparing individuals to the manifold population, which combines a mapping to the manifold and the path along the manifold to reach its origin (Fig. 1a). The originality of our method resides in the use of motion abnormality maps as input, as introduced in [2], which allows the gradation of the disease and the definition of a physiologically meaningful origin within the manifold, representing a normal motion pattern. Each pathological pattern is therefore considered a deviation from normality along a manifold structure. We present the application of the proposed method to septal flash (SF), a specific motion pattern of intra-ventricular dyssynchrony associated to a high response rate to CRT [11].

2 Methods

The computation of a distance between individuals and a given population considered as a pathologic deviation from normality consists of two steps: the quantification of abnormality for each subject in the dataset, and the estimation of a manifold for this population, constrained to pass by an origin representing normality.

2.1 Atlas-Based Maps of Motion Abnormalities

The inputs for our method consist of 2D spatiotemporal maps of myocardial motion abnormalities, obtained from a statistical atlas of motion built from healthy volunteers [2]. Each map corresponds to one subject in the dataset, and is used as a 2D image in which the horizontal dimension is time (systole) and the vertical one is the position along the septum. Each pixel value corresponds to a p -value index used to locally encode abnormality, in a logarithmic scale, multiplied by the sign of the radial velocity. This choice was made to highlight the inward and outward events of SF, when present (Fig 2a). The color-code associates blue and red color to highly abnormal inward and outward motion of the septum, respectively. According to these conventions, the origin used to constrain the manifold (Sec. 2.2) is defined as an image having 0 value at each pixel, representing a normal motion pattern.

2.2 Manifold-Based Distances to a Population

All the images considered in this paper belong to an ambient space \mathcal{A} . Let's denote $\mathcal{I} = \{\mathbf{I}_0, \dots, \mathbf{I}_N\} \subset \mathcal{A}$ the dataset of $N + 1$ images used for the manifold estimation. The image \mathbf{I}_0 corresponds to the manifold origin for normality. This image is added to the original dataset $\{\mathbf{I}_1, \dots, \mathbf{I}_N\}$ before any computation, so that every image \mathbf{I}_i , $i > 0$ is connected to \mathbf{I}_0 through the isomap graph resulting from the computations described below. This amounts to considering every element of \mathcal{I} as a deviation from the origin along a specific path on the manifold structure.

The space of manifold coordinates is denoted $\mathcal{C} \subset \mathbb{R}^m$, m being the dimensionality of the manifold, while $f : \mathcal{A} \rightarrow \mathcal{C}$ and $g : \mathcal{C} \rightarrow \mathcal{A}$ stand for the correspondence functions between \mathcal{A} and \mathcal{C} . The computation of these functions is based on interpolation techniques adapted from [1] and explained in the following paragraphs. We denote $d : \mathcal{A} \rightarrow \mathbb{R}$ the metric used to compare elements of \mathcal{A} .

Manifold estimation. The isomap algorithm [13] is used to estimate the manifold. First, a graph is built for the dataset \mathcal{I} , based on the k -NN algorithm, connecting all the images among themselves according to the metric d . Then, Euclidean embedding of the manifold data provides a set of coordinates $\mathcal{X} = \{\mathbf{x}_0, \dots, \mathbf{x}_N\} \subset \mathcal{C}$.

From ambient space to manifold coordinates. The estimation of $f : \mathcal{A} \rightarrow \mathcal{C}$ can be formulated as an exact matching problem on a reproducible kernel Hilbert space \mathcal{V} [12] of functions $\mathcal{A} \rightarrow \mathcal{C}$, namely:

$$\operatorname{argmin}_{f \in \mathcal{V}} \left(\frac{1}{2} \|f\|_{\mathcal{V}}^2 \right) \quad \text{under the constraint } f(\mathbf{I}_i) = \mathbf{x}_i, \quad \forall i \in [0, N] \quad (1)$$

with solution: $f(\mathbf{I}) = \sum_{i=0}^N K_f(\mathbf{I}, \mathbf{I}_i) \cdot \mathbf{a}_i$ with $\mathbf{a}_{\mathcal{I}} := \mathbf{K}_f^{-1} \cdot \mathbf{x}_{\mathcal{I}}$,

where \mathbf{K}_f is the matrix $(K_f(\mathbf{I}_i, \mathbf{I}_j))_{(i,j) \in [0,N]^2}$, K_f being chosen of the exponential form $K_f(\mathbf{I}, \mathbf{J}) := \exp(-d(\mathbf{I}, \mathbf{J})^2/\sigma_f^2)$, $(\mathbf{I}, \mathbf{J}) \in \mathcal{A}^2$, σ_f being its bandwidth, and $\mathbf{a}_{\mathcal{I}}$ and $\mathbf{x}_{\mathcal{I}}$ the vectors $(\mathbf{a}_i)_{i \in [0,N]}$ and $(\mathbf{x}_i)_{i \in [0,N]}$, respectively.

Back from manifold coordinates to the ambient space. The estimation of $g: \mathcal{C} \rightarrow \mathcal{A}$ is a variant of the previous computation, formulated as an inexact matching problem on a reproducible kernel Hilbert space \mathcal{W} [12] of functions $\mathcal{C} \rightarrow \mathcal{A}$, with a constraint to force the manifold to pass by the origin image \mathbf{I}_0 :

$$\operatorname{argmin}_{g \in \mathcal{W}} \left(\frac{1}{2} \|g\|_{\mathcal{W}}^2 + \frac{\gamma}{2} \sum_{i=1}^N d(g(\mathbf{x}_i), \mathbf{I}_i) \right) \quad \text{under the constraint } g(\mathbf{x}_0) = \mathbf{I}_0 \quad (2)$$

with solution: $g(\mathbf{x}) = \sum_{i=0}^N K_g(\mathbf{x}, \mathbf{x}_i) \cdot \mathbf{b}_i$ with $\mathbf{b}_{\mathcal{I}} := (\mathbf{K}_g + \frac{1}{\gamma} \mathbf{M})^{-1} \cdot \mathbf{I}_{\mathcal{I}}$,

where \mathbf{K}_g is the matrix $(K_g(\mathbf{x}_i, \mathbf{x}_j))_{(i,j) \in [0,N]^2}$, K_g being chosen of the exponential form $K_g(\mathbf{x}, \mathbf{y}) := \exp(-\|\mathbf{x} - \mathbf{y}\|^2/\sigma_g^2)$, $(\mathbf{x}, \mathbf{y}) \in \mathcal{C}^2$, σ_g being its bandwidth, \mathbf{M} is the matrix $(M_{i,j})_{(i,j) \in [0,N]^2}$, with $M_{i,i} = 1 \forall i \neq 0$ and 0 otherwise, and $\mathbf{b}_{\mathcal{I}}$ and $\mathbf{I}_{\mathcal{I}}$ the vectors $(\mathbf{b}_i)_{i \in [0,N]}$ and $(\mathbf{I}_i)_{i \in [0,N]}$, respectively.

The addition of such a constraint is illustrated in Fig. 1b, which displays the interpolated curve obtained from a 1D synthetic dataset using inexact matching before and after forcing the curve to pass by one point, as described in Eq. 2.

Mapping to the manifold and induced distance. Any image $\mathbf{I} \in \mathcal{A}$ can be associated to an element of the manifold $\hat{\mathbf{I}}$ by means of the composition of the above-defined functions, using $\hat{\mathbf{I}} = g(f(\mathbf{I}))$. This composition allows defining a distance between any image $\mathbf{I} \in \mathcal{A}$ and the manifold [6], namely: $d_{\text{mapping}}(\mathbf{I}) = d(\hat{\mathbf{I}}, \mathbf{I})$. This distance is complemented by a second one, which compares individuals to normality along the manifold structure: $d_{\text{manifold}}(\mathbf{I}) = \|f(\mathbf{I}) - \mathbf{x}_0\|^2$. Total distance to normality is then written as $\sqrt{(d_{\text{mapping}})^2 + (d_{\text{manifold}})^2}$.

3 Experiments and Results

Patient population and processed data. Using the method presented in Sec. 2.2, a manifold was estimated from a population of 50 CRT candidates with SF. This manifold is expected to represent pathologic deviations from normality, each point of the manifold being a SF pattern. A second dataset was used for testing the distances proposed in Sec. 2.2. This population was made of 38 CRT candidates (7 having SF and 31 without SF) and 21 healthy volunteers. All patient data was acquired before the implantation of the CRT device. The presence of SF was assessed by two experienced cardiologists, from the visual inspection of echocardiographic M-mode images, as described in [11].

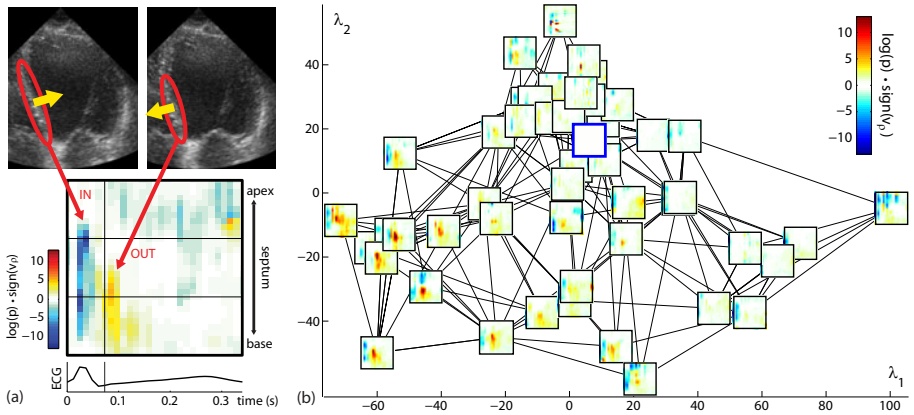


Fig. 2. (a) Map of septal motion abnormalities for one CRT candidate with SF. The color-scale encodes abnormality (p -value) in a logarithmic scale, multiplied by the sign of the radial velocity \mathbf{v}_ρ to highlight SF (Sec. 2.1). (b) Two-dimensional embedding of the manifold of SF p -value maps (output of isomap) according to its two first dimensions. The blue-framed map corresponds to the image used to constrain the manifold, representing a normal motion pattern.

A 2D spatiotemporal map of myocardial motion abnormalities obtained from a statistical atlas of motion [2] was associated to each subject, as explained in Sec. 2.1. The atlas was built from the set of 21 healthy volunteers. The abnormality maps had a size of 31×20 pixels, corresponding to the sampling of the systolic period (horizontal dimension) and the septum along its medial line (vertical dimension), respectively.

The sum of squared differences was used for the metric $d : \mathcal{A} \rightarrow \mathbb{R}$. The number of neighbors for the k -NN computations was set to $k = 5$, which guaranteed that all the images from the manifold dataset were connected among themselves, as tested experimentally. A two-dimensional embedding of the computed manifold (output of isomap) is represented in Fig. 2b, showing the link between each image and its k -NN. The bandwidths for the kernels K_f and K_g introduced in Sec. 2.2 were set to the average k -NN distances over the manifold population and its corresponding set of coordinates \mathcal{X} , respectively. The value of γ involved in the inexact matching problem (Eq. 2) was set to 10, representing a compromise between the smoothness of the manifold and its closeness to the data.

Manifold accuracy. The influence of the manifold dimensionality on the reconstruction error was estimated by computing the average and standard deviation of $d_{mapping}$ over the manifold population, as shown in Fig. 3a. The curve reaches 95% of its final value when the manifold dimensionality is higher than 10, which is the value we chose for the rest of the paper. The reconstruction error obtained with a linear approximation of the manifold dataset (using PCA) is higher, as visible in this figure, justifying the choice of nonlinear techniques to characterize the population of CRT candidates with SF.

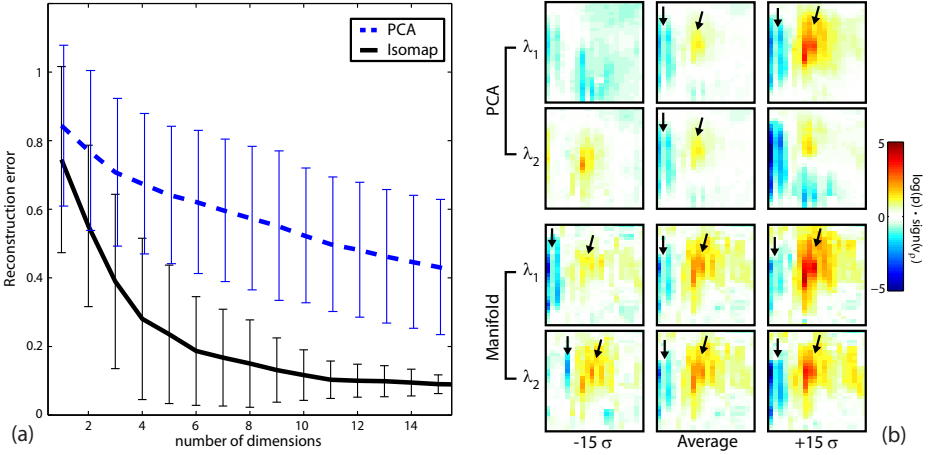


Fig. 3. (a) Evolution of the reconstruction error against the number of dimensions used. Comparison between PCA (dashed blue) and isomap (black). Values correspond to the average \pm standard deviation over the manifold dataset \mathcal{I} . (b) Variations around the average map along the two first principal directions of the manifold dataset \mathcal{I} , obtained using either PCA or isomap. Arrows indicate the inward and outward events of SF, when this pattern is present on the map.

The limitations of PCA on the studied dataset, compared to isomap, are also visible on Fig. 3b. This figure represents the variations around the average map along the two first principal directions of the manifold dataset \mathcal{I} , obtained using either PCA or isomap. As indicated by the black arrows, PCA does not guarantee that the computed maps still correspond to a SF, while this pattern is preserved by the use of manifold-learning.

Distance to the manifold. Figure 4 represents the distance between all the subjects involved in this study and the manifold. We separated the analysis between $d_{mapping}$ and $d_{manifold}$ for interpretation purposes. The patients from the manifold dataset have low $d_{mapping}$, which corresponds to the reconstruction error plotted in Fig. 3a. As the manifold is built from this population, they largely span the space associated to $d_{manifold}$. There is no SF patient from the manifold dataset close to the origin according to $d_{manifold}$, while almost all the healthy volunteers have lower values (vertical lines indicate the median and 1st/3rd quartiles of $d_{manifold}$ for the healthy subjects). This provides an estimation of the threshold above which SF can be detected, as being a deterioration from normality. This threshold may come from the accuracy of the patient selection process using M-mode images [11], and from the minimum accuracy of the abnormality maps [2]. Among the testing subjects, patients having SF are closer to the manifold than patients without SF, according to $d_{mapping}$. Higher values of $d_{manifold}$ are observed in the subjects having higher SF abnormalities on the maps. A larger bandwidth

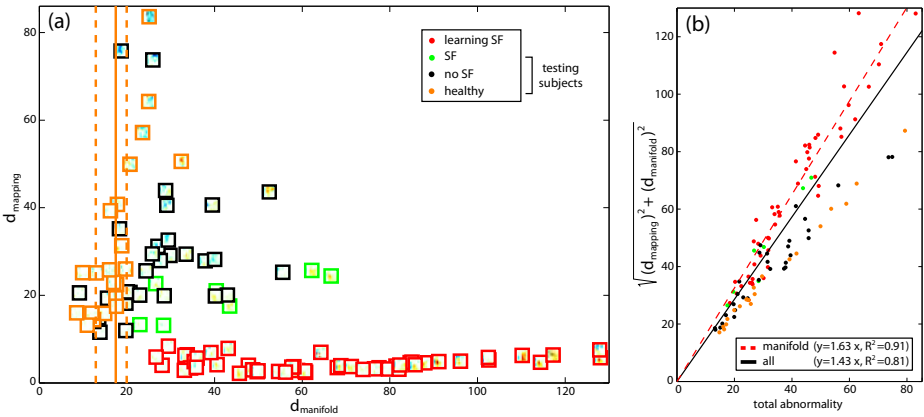


Fig. 4. (a) Subject ordering according to $d_{manifold}$ and $d_{mapping}$, used as horizontal and vertical axis, respectively. Vertical orange lines indicate the median and $1^{st}/3^{rd}$ quartiles of $d_{manifold}$ for the healthy subjects. (b) Patient ordering according to the amount of total abnormality and the proposed distance $\sqrt{(d_{mapping})^2 + (d_{manifold})^2}$.

for the kernels K_f and K_g would bring the testing patients with SF closer to the manifold, but would also increase the reconstruction error.

As the 2D maps processed in this study locally contain a measure of abnormality, it is also of interest to compare the total abnormality of each map against the distances introduced in Sec. 2.2. This comparison is shown in Fig. 4b. Total abnormality was computed for each subject using the L^2 norm of its abnormality map. Linear regression over the plotted data led to R^2 coefficients of 0.91 (manifold data only, dashed red line) and 0.81 (whole data, black line). This suggests that the dimensionality reduction inherent to the manifold estimation preserves the concept of abnormality embedded in the processed maps.

4 Conclusion

We have proposed a method for modeling a specific pathological motion pattern as a manifold. This manifold represents pathological motion as a deviation from normality, being by construction the manifold origin. The method was used to compute a distance between individuals and a given pathological pattern. Experiments demonstrate the need of nonlinear embedding of the learning data, and the relevance of the proposed method for grading different stages of motion abnormality. In the context of CRT, the method can improve the selection of responders to the therapy, allowing reproducible comparison of a new candidate to specific patterns of mechanical dyssynchrony that condition CRT outcome.

Acknowledgments. This work was supported by the Spanish Industrial and Technological Development Center (cvREMOD CEN-20091044), the Spanish

Ministry of Science and Innovation, Plan E and ERDF (STIMATH TIN2009-14536-C02-01), and the European Commission's Seventh Framework Program (EuHeart FP7-ICT-224495). AFF is partially funded by the ICREA-Academia program.

References

1. Davis, B.C., Fletcher, P.T., Bullitt, E., Joshi, S.: Population shape regression from random design data. *Int. J. Comput. Vis.* 90, 255–266 (2010)
2. Duchateau, N., De Craene, M., Piella, G., Silva, E., Doltra, A., Sitges, M., Bijmens, B.H., Frangi, A.F.: A spatiotemporal statistical atlas of motion for the quantification of abnormal myocardial tissue velocities. *Med. Image Anal.* 15(3), 316–328 (2011)
3. Etyngier, P., Keriven, R., Segonne, F.: Projection onto a shape manifold for image segmentation with prior. In: 14th IEEE International Conference on Image Processing, pp. IV361–IV364. IEEE Press, New York (2007)
4. Fletcher, P.T., Lu, C., Pizer, S.M., Joshi, S.: Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE Trans. Med. Imaging* 23(8), 995–1005 (2004)
5. Fornwalt, B.K.: The dyssynchrony in predicting response to cardiac resynchronization therapy: A call for change. *J. Am. Soc. Echocardiogr.* 24(2), 180–184 (2011)
6. Gerber, S., Tasdizen, T., Fletcher, P.T., Joshi, S., Whitaker, R.: Manifold modeling for brain population analysis. *Med. Image Anal.* 14(5), 643–653 (2010)
7. Kim, K.-H., Choi, S.: Neighbor search with global geometry: a minimax message passing algorithm. In: 24th International Conference on Machine Learning, pp. 401–408 (2007)
8. Kwok, J.T.-Y., Tsang, I.W.-H.: The pre-image problem in kernel methods. *IEEE Trans. Neural Netw.* 15(6), 1517–1525 (2004)
9. Martinez, A.M., Kak, A.C.: PCA versus LDA. *IEEE Trans. Pattern Anal. Mach. Intell.* 23(2), 228–233 (2001)
10. Mika, S., Schölkopf, B., Smola, A., Müller, K.R., Scholz, M., Rätsch, G.: Kernel PCA and de-noising in feature spaces. *Adv. Neural Inf. Process Syst.* 11, 536–542 (1999)
11. Parsai, C., Bijmens, B.H., Sutherland, G.R., Baltabaeva, A., Claus, P., Marciniak, M., Paul, V., Scheffer, M., Donal, E., Derumeaux, G., Anderson, L.: Toward understanding response to cardiac resynchronization therapy: left ventricular dyssynchrony is only one of multiple mechanisms. *Eur. Heart J.* 30(8), 940–949 (2009)
12. Saitoh, S.: *Theory of Reproducing Kernels and its Applications*. Pitman Res. Notes Math. Ser., p. 189. Wiley, Chichester (1988)
13. Tenenbaum, J.B., De Silva, V., Langford, J.C.: A global geometric framework for nonlinear dimensionality reduction. *Science* 290(5500), 2319–2323 (2000)
14. Worsley, K.J., Taylor, J.E., Tomaiuolo, F., Lerch, J.: Unified univariate and multivariate random field theory. *NeuroImage* 23(suppl. 1), S189–S195 (2004)
15. Zheng, W.S., Lai, J.-H., Yuen, P.C.: Penalized preimage learning in kernel principal component analysis. *IEEE Trans. Neural Netw.* 21(4), 551–570 (2010)

Detecting Outlying Subjects in High-Dimensional Neuroimaging Datasets with Regularized Minimum Covariance Determinant

Virgile Fritsch^{1,2}, Gael Varoquaux^{3,1,2}, Benjamin Thyreau²,
Jean-Baptiste Poline^{2,1}, and Bertrand Thirion^{1,2}

¹ Parietal Team, INRIA Saclay-Île-de-France, Saclay, France
virgile.fritsch@inria.fr
<http://parietal.saclay.inria.fr>

² CEA, DSV, I²BM, Neurospin bât 145, 91191 Gif-Sur-Yvette, France

³ Inserm, U992, Neurospin bât 145, 91191 Gif-Sur-Yvette, France

Abstract. Medical imaging datasets used in clinical studies or basic research often comprise highly variable multi-subject data. Statistically-controlled inclusion of a subject in a group study, i.e. deciding whether its images should be considered as samples from a given population or whether they should be rejected as outlier data, is a challenging issue. While the informal approaches often used do not provide any statistical assessment that a given dataset is indeed an outlier, traditional statistical procedures are not well-suited to the noisy, high-dimensional, settings encountered in medical imaging, *e.g.* with functional brain images. In this work, we modify the classical *Minimum Covariance Determinant* approach by adding a regularization term, that ensures that the estimation is well-posed in high-dimensional settings and in the presence of many outliers. We show on simulated and real data that outliers can be detected satisfactorily, even in situations where the number of dimensions of the data exceeds the number of observations.

Keywords: Outlier detection, Minimum Covariance Determinant, regularization, robust estimation, neuroimaging, fMRI.

1 Introduction

Between-subject variability is a prominent effect in many fields of medical imaging, and particularly in brain imaging. While part of this variability can be viewed as normal fluctuations within a population or across repeated measurements, and can be considered as an effect of interest for diagnosis problems, part of it may be a confound, related to scanner instabilities, experimental issues, or acquisition artifacts. Such confounding factors can be much larger than the effects of interest: for instance, in functional neuroimaging, the variability related to acquisition issues (motion, defective experimental setup, scanner spikes) can mask the true effect of interest, which is the variability in brain functional organization related to diseases, psychological or genetic factors.

The detection of abnormal data, or outlier detection, is important in order to ensure that the ensuing statistical analysis will be robust to such undesired effects. This detection should be automated for the sake of reproducibility and to be time efficient, as cohorts can now encompass up to several hundreds of subjects. This detection is challenging because *i)* images, in particular brain images, are *complex, high-dimensional objects* with some unknown latent structure; *ii)* the problem is *unsupervised*, in the sense that outlier detection procedures can in general not be calibrated on training data; and *iii)* in many cases, it is impossible to *normalize* the signal or its variability.

So far, high-dimensional analysis procedures have been confined to high SNR data, such as anatomical images, e.g. with the use of manifold learning techniques [13]. These, however, are not robust to outlier data, and are not applicable to functional Magnetic Resonance Imaging (fMRI) since they may easily be confounded by noise. As a first step to alleviate this issue, univariate outlier detection methods have been proposed for fMRI, in which one particular image feature is studied, and compared to other data [6,12]. Kherif *et al.* [5] point out the need of homogeneous datasets in fMRI studies and propose a model-based multivariate framework as a solution. However, their work is restricted to small cohorts and does not discuss statistical control.

While the robust statistics literature generally considers that problems with a number of dimensions comparable to the number of observations cannot be addressed in model-based approaches, we investigate whether outlier detection is still possible in that setting. Specifically, we modify the Minimum Covariance Determinant method [8] so that its performance approaches the level of non-parametric methods, such as one-class Support Vector Classification [2]. We describe the new robust estimator in the next section and show its well-posedness. We then perform some experiments on simulated data and assess the behaviour of the proposed method with respect to state-of-the-art techniques. Finally, we describe the application of our approach to an fMRI dataset, where we show that outliers can still be detected on medium-sized groups of subjects.

2 Robust Location and Covariance Estimates

We focus on a model-based approach, as it yields more interpretable results as well as a probabilistic control of false detections: Assuming a high dimensional Gaussian model, an observation $x_i \in \mathbb{R}^p$ within a set X can be characterized as outlier whenever it has a large Mahalanobis distance to the mean of the data distribution, defined as $d_{\hat{\mu}, \hat{\Sigma}}^2(x_i) = (x_i - \hat{\mu})^T \hat{\Sigma}^{-1}(x_i - \hat{\mu})$, $\hat{\mu}$ and $\hat{\Sigma}$ being respectively estimates of the dataset location and covariance. Crucially, robust estimators of location and covariance have to be used for the computation of these distances.

MCD estimator and FastMCD algorithm. The state-of-the-art robust covariance estimator for multidimensional Gaussian data is Rousseeuw’s Minimum Covariance Determinant (MCD) estimator [8], which can be computed using the Fast-MCD algorithm [10]. Given a dataset with n p -dimensional observations, MCD

aims at finding h observations (referred to as the *support*), the scatter matrix of which has a minimal determinant. For the scatter matrix to be well-conditioned, h must be greater than $h_{\min} = \frac{n+p+1}{2}$. As $\frac{p}{n}$ becomes large, h_{\min} increases so outliers are potentially included in the covariance estimation if there are more than $\frac{n-p-1}{2}$ of them. When $p = n - 1$, the MCD estimator is equivalent to the unbiased maximum likelihood estimator, which is not robust. Finally, if $p \geq n$, the MCD estimator is not defined. To alleviate these issues we propose to use half of the observations in the support ($h = \frac{n}{2}$) and compensate the shortage of data for covariance estimation with regularization.

Regularized MCD estimator (R-MCD). We consider ridge regularization: let $\lambda \in \mathbb{R}^+$ be the amount of regularization, (μ_r, Σ_r) the location and covariance estimates of a $n \times p$ dataset X that maximize the penalized negative log-likelihood:

$$(\mu_r, \Sigma_r) = \operatorname{argmin}_{\mu, \Sigma} \left(\log |\Sigma| + \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^T \Sigma^{-1} (x_i - \mu) + \lambda \operatorname{Tr} \Sigma^{-1} \right), \quad (1)$$

yielding $\Sigma_r = \frac{X^T X}{n-1} + \lambda \operatorname{Id}_p$ and $\mu_r = \frac{1}{n} X^T \mathbf{1}_n$.

Convergence of the Fast R-MCD algorithm. Fast-MCD is an iterative algorithm that successively culls out outliers using Mahalanobis distances defined with the covariance of the most homogeneous fraction of the data. In our new algorithm, *Fast-R-MCD*, we replace the sample covariance matrix used in MCD to define the Mahalanobis distance by the ridge estimate. The convergence of Fast-R-MCD stems from the following lemma, that generalizes the proof of convergence of Fast-MCD [4]:

$$\forall \eta > 0, \quad (\mu_r, \Sigma_r) = \begin{cases} \operatorname{argmin}_{\mu, \Sigma} |\Sigma|, \\ \text{s.t. } \mathbb{E} [(X - \mu)^T \Sigma^{-1} (X - \mu)] + \lambda \operatorname{tr} \Sigma^{-1} = \eta \end{cases} \quad (2)$$

which straightforwardly implies that the determinant of Σ_r will decrease at each iteration of the Fast-R-MCD algorithm.

Setting the regularization parameter λ . Starting with an initial guess for $\lambda = \frac{\operatorname{tr}(\hat{\Sigma})}{np}$ where $\hat{\Sigma}$ is the unbiased empirical covariance matrix of the whole dataset, we isolate an uncontaminated set of $\frac{n}{2}$ observations, as in the Fast-MCD approach. Let $\lambda = \delta \frac{\operatorname{tr}(\hat{\Sigma}_{\text{pure}})}{np}$, where $\hat{\Sigma}_{\text{pure}}$ is the empirical covariance matrix of the uncontaminated dataset. We choose δ so as to maximize the ten-fold cross-validated log-likelihood of the uncontaminated dataset.

3 Experiments

We compared the outlier detection accuracy that can be obtained from the Mahalanobis distances of the samples, using respectively MCD and R-MCD.

3.1 Simulations

Data generation. In our simulations, we sample a core set of $n - q$ ($q < \frac{n}{2}$) observations from a $\mathcal{N}(0_p, \Sigma)$ distribution corresponding to regular observations (also called *inliers*). We add q outliers from a $\mathcal{N}(\mu_q, \Sigma_q)$ distribution ($\mu_q \in \mathbb{R}^p, \Sigma_q \in S_n^+(p)$), thus generating a total of n observations with p features. We use three outliers types (see Fig. 1):

Variance outliers are obtained by setting $\Sigma_q = \alpha \Sigma, \alpha > 1$ and $\mu_q = 0_p$. This situation models signal normalization issues, where the amount of variance in outlier observations is abnormally large.

Multi-modal outliers are obtained by setting $\Sigma_q = \Sigma$ and $\mu_q \neq 0_{\mathbb{R}^p}$, which simulates the presence of an heterogeneous population.

Multivariate outliers are obtained by setting $\mu_q = 0_p, \Sigma_q = \Sigma + \alpha \mathbf{a} \mathbf{a}^T$ where $\mathbf{a} = \frac{\mathbf{a}_{\text{rand}}}{\|\mathbf{a}_{\text{rand}}\|_2}$ and \mathbf{a}_{rand} is a vector p -dimensional vector with coordinates drawn from a Bernoulli distribution $\mathcal{B}(\frac{1}{2})$. This model simulates outliers as sets of points having potentially abnormally high values in some random directions.

In our experiments, we also investigated the influence of Σ 's condition number $\kappa(\Sigma) = \|\Sigma\|_2 \cdot \|\Sigma^{-1}\|_2$ and contamination rate $\frac{q}{n}$.

Methods comparison. Given a simulated dataset, we estimated the location and covariance of the data using MCD and R-MCD estimators. Both were computed with the Fast-MCD (or Fast-R-MCD) algorithm without consistency and re-weighting steps (see [10]), leading to what we call *raw estimates*. The parameter that influences the most the relative performances of MCD- and R-MCD-based outlier detection methods is the $\frac{p}{n}$ ratio. Every other parameter being fixed, we averaged 100 ROC curves for each value of $\frac{p}{n}$ in a given range, and finally expressed Area Under Curve (AUC) as a function of $\frac{p}{n}$.

We also compare the R-MCD sensitivity with the One-class SVM sensitivity, holding the latter as a reference since it is not limited by any prior shape of the separation between in- and outlying observations. We used a *RBF* kernel and selected its bandwidth γ with an heuristic inspired by [11]: $\gamma = \frac{0.01}{\Delta}$, where Δ is the 10th percentile of the pairwise distances histogram of the observations.

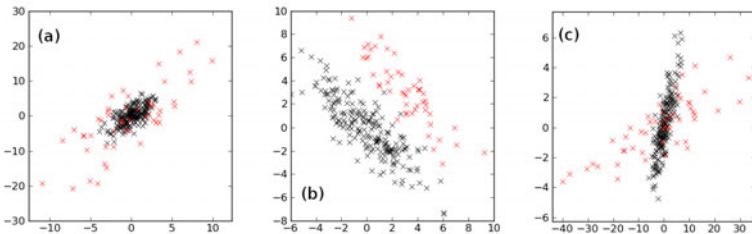


Fig. 1. Three different ways to generate multivariate outliers for Gaussian data. **(a)** all directions ($\alpha = 3$). **(b)** second cluster ($\mu_q = 3 \times \mathbf{1}_{\mathbb{R}^2}$). **(c)** multivariate ($\alpha = 5$). Outliers are represented in red and inliers in black. $\kappa(\Sigma) = 10$. Contamination is 40%.

3.2 Outliers Identification in Functional Neuroimaging

We used data from a large functional neuroimaging database containing several fMRI contrast images in more than 1500 subjects. 3T scanners from multiple manufacturers were used for acquiring the data with TR = 2200 ms, TE = 30 ms, and flip angle = 75°. Standard preprocessing were performed on the data using the SPM8 software. Here we focus on a control contrast that shows brain regions implied in auditory tasks as opposed to visual tasks.

We used a probabilistic brain atlas [7] to extract an average activation intensity value from 145 regions of interest in all the contrast images. We then performed an initial outlier detection at $P < 0.1$ familywise corrected, including more than 1000 subjects. With such a small $\frac{p}{n}$ value, a statistically controlled outlier detection could be done using the MCD estimate. The outliers list obtained from this first outlier detection was then held as a ground truth for further outlier detection experiments performed on reduced sample, using MCD and R-MCD estimators. Note that for very small samples, we could not use the MCD-based outliers detection method. The outliers lists were compared to the ground truth and ROC curves were hence constructed. For each sample size, we repeated the detection 10 times with 10 different, randomly selected samples.

4 Results

4.1 Simulation Results

We first give the results on a 30-dimensional dataset with a 40% contamination rate ($\frac{p}{n} = 0.4$), generated from the multivariate outliers model. We show the case where $\kappa(\Sigma) = 1000$. The accuracy of the R-MCD-based method is much higher than the accuracy of the MCD-based method as soon as $\frac{p}{n} > 0.3$ (Fig 2). Using regularization, it is possible to go beyond the $p = n$ limit, keeping an AUC greater than 0.70.

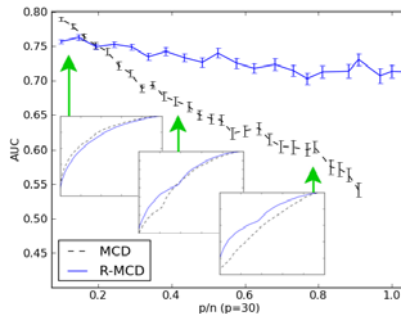


Fig. 2. AUC for MCD- and R-MCD-based outliers detection methods. 40% multivariate outliers are generated ($\alpha = 5$, $\kappa(\Sigma) = 1000$). The R-MCD-based method keeps an AUC of $\simeq 0.70$ up to $\frac{p}{n} = 3.5$ (not shown) while the MCD-based method breaks down.

Outliers type. Table 1a summarizes the AUC results for experiments with the variance outliers model. In this case, the MCD performance drops when $\frac{p}{n} > 0.5$ while the R-MCD-based method always achieves at least a 80% accuracy. In the multimodal case (Table 1b), the R-MCD-based method is also more successful at detecting outliers.

Covariance matrix condition number and contamination rate. The methods' performance depends weakly on the condition number. A small condition number yields better results with R-MCD while a high condition number gives advantage to the use of the MCD estimator when $\frac{p}{n}$ is small. The higher the contamination rate, the more MCD estimator is likely to break. On the other hand, the performance of a R-MCD-based detection method is stable (Table 2).

Comparison to One-class SVM. In both cases of variance and multivariate outliers, One-class SVM achieves a better specificity/sensitivity compromise than R-MCD-based outlier detection method. Yet, for a $\frac{p}{n}$ ratio of the order of .5, the R-MCD performance remains comparable to that of the One-class SVM, with an asymptotic score that remains below (0.05 difference in the AUC).

4.2 Application on a Real Dataset

We give the averaged ROC curves for detecting outliers on an auditory task in Fig 4. The reference outlier detection was performed on 1118 subjects, each being described by 145 features. The results shown correspond to 10 random sets of 290 subjects. In the useful range ($FP \leq 5\%$), R-MCD outperforms MCD. Even with only 100 samples ($\frac{p}{n} = 1.5$), R-MCD can still be used to find outliers, as Fig 4 and Fig 5 demonstrate. On this latter figure, outlying subjects 1, 2 and 3 indeed exhibit much variable activity patterns than subjects A or B, despite the presence of a few mistakes (subject 4).

5 Discussion

In high-dimensional Gaussian datasets, our results show that Regularized MCD can reach a significantly higher sensitivity in outlier detection than the standard MCD. We assumed that neuroimaging data are distributed according to a multivariate Gaussian distribution. This strong hypothesis lead us to focus on Mahalanobis-distances-based approaches since they can exploit the assumed shape of the dataset to *estimate its covariance matrix*. Since R-MCD systematically deals with half of the observations, it is not subject to the known masking and swamping effects [9]. We plan to investigate a ℓ_1 norm for covariance regularization, as it may fit with standard hypotheses on brain covariance structure.

Under the Gaussian assumption we made, outlier detection with the R-MCD estimator is the only method so far that both holds in high-dimension and allows a probabilistic control of the false detection rate. Although the One-class SVM non-parametric algorithm achieves a better sensitivity/specificity compromise and is still applicable with non-Gaussian data, its lack of interpretability

Table 1. AUC values: **a: left** variance outliers model ($p = 30, q/n = 40\%, \alpha = 1.25$), **b: right** multimodal outliers model ($p = 30, q/n = 20\%, \mu_q = 2 \cdot \mathbf{1}_p$)

p/n	0.1	0.2	0.3	0.4	0.5	0.7	0.8
MCD	0.86	0.82	0.77	0.73	0.70	0.66	0.63
R-MCD	0.87	0.86	0.85	0.85	0.84	0.82	0.82

p/n	0.1	0.2	0.3	0.4	0.5	0.7	0.8
MCD	0.62	0.60	0.58	0.57	0.55	0.55	0.51
R-MCD	0.76	0.77	0.78	0.81	0.78	0.75	0.77

Table 2. Influence of the contamination rate. $p = 30$, multivariate outliers ($\alpha = 5$). Unlike MCD, the R-MCD performances are independent of the contamination value.

q/n	10 %				20 %				30 %				40 %			
p/n	0.1	0.3	0.6	0.9	0.1	0.3	0.6	0.9	0.1	0.3	0.6	0.9	0.1	0.3	0.6	0.9
MCD	0.83	0.77	0.68	0.56	0.82	0.74	0.65	0.57	0.81	0.72	0.65	0.55	0.79	0.71	0.64	0.54
R-MCD	0.77	0.74	0.72	0.68	0.76	0.75	0.72	0.71	0.76	0.74	0.72	0.72	0.76	0.75	0.74	0.71

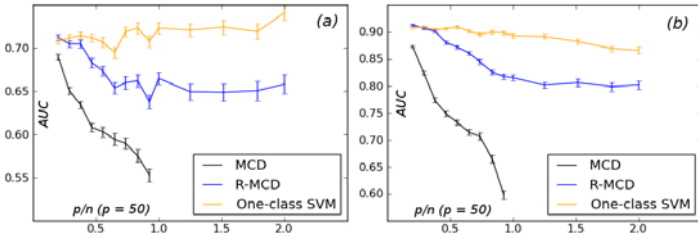


Fig. 3. One-class SVM comparison with 40% contamination and $\kappa(\Sigma) = 100$. (a) AUC for multivariate outliers ($\alpha = 5$). (b) AUC for variance outliers ($\alpha = 1.25$).

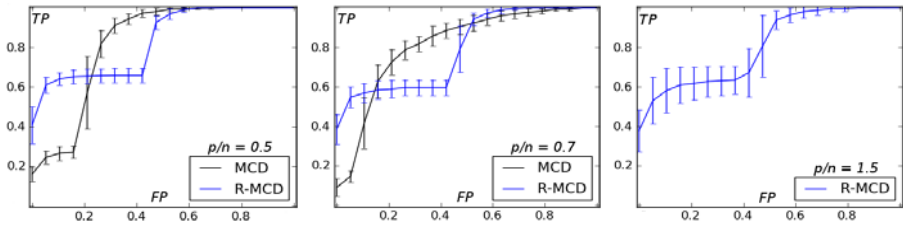


Fig. 4. ROC curves showing that R-MCD outperforms MCD on real fMRI data

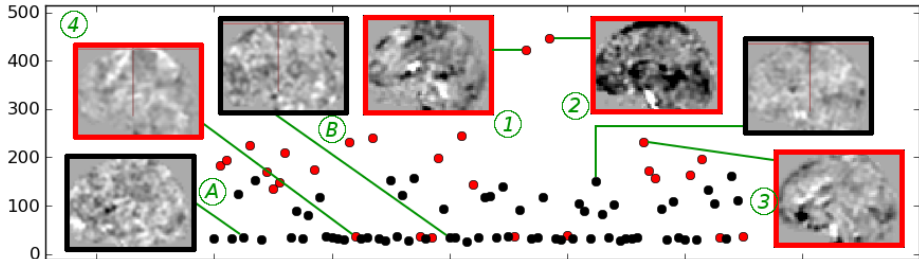


Fig. 5. R-MCD-based Mahalanobis distances of a small sample. The higher the Mahalanobis distance, the higher the probability for an observation to be tagged as outlying. Points in red are outliers subjects according to the whole population.

and statistical control, as well as the difficulty to tune its parameters, makes it unsuitable in a medical context.

6 Conclusion

We introduced the R-MCD estimator, a regularized version of a robust covariance estimator commonly used to detect outlying observations on the basis of their Mahalanobis distances. We showed that the Fast-MCD algorithm is still valid to compute this new estimator. Our application to neuroimaging, where studies have a high exclusion rate, shows that it is possible to build automatic procedures to detect outliers even though the number of descriptors is higher than the number of available subjects. This property is of broad interest in medical applications where heterogeneous populations have to be considered and relies on an objective assessment of normal variability.

This work was supported by a Digiteo DIM-Lsc grant (HiDiNim project, N°2010-42D). JBP was partly funded by the IMAGEN project, which receives research funding from the E.U. Community's FP6, LSHM-CT-2007-037286. This manuscript reflects only the author's views and the Community is not liable for any use that may be made of the information contained therein.

References

1. Aljabar, P., Wolz, R., Srinivasan, L., Counsell, S., Boardman, J.P., Murgasova, M., Doria, V., Rutherford, M.A., Edwards, A.D., Hajnal, J.V., Rueckert, D.: Combining morphological information in a manifold learning framework: application to neonatal MRI. *Med. Image Comput. Comput. Assist. Interv.* 13, 1 (2010)
2. Gardner, A., Krieger, A., Vachtsevanos, G., Litt, B.: One-class novelty detection for seizure analysis from intracranial EEG. *J. Mach. Learn. Res.* 7, 1025 (2006)
3. Gerber, S., Tasdizen, T., Joshi, S., Whitaker, R.: On the manifold structure of the space of brain images. *Med. Image Comput. Comput. Assist. Interv.* 12, 305 (2009)
4. Grubel, R.: A minimal characterization of the covariance matrix. *Metrika* 35, 49 (1988)
5. Kherif, F., Flandin, G., Ciuciu, P., Benali, H., Simon, O., Poline, J.B.: Model based spatial and temporal similarity measures between series of functional magnetic resonance images. *Med. Image Comput. Comput. Assist. Interv.*, 509 (2002)
6. Penny, W.D., Kilner, J., Blankenburg, F.: Robust bayesian general linear models. *Neuroimage* 36, 661 (2007)
7. Perrot, M., Rivière, D., Tucholka, A., Mangin, J.F.: Joint bayesian cortical sulci recognition and spatial normalization. *Inf. Process. Med. Imaging* 21, 176–187 (2009)
8. Rousseeuw, P.J.: Least median of squares regression. *J. Am. Stat. Ass.* 79, 871 (1984)
9. Rousseeuw, P.J., Hubert, M.: Robust statistics for outlier detection. *WIREs Data Mining Knowl. Discov.* 1, 73 (2011)
10. Rousseeuw, P.J., Van Driessen, K.: A fast algorithm for the minimum covariance determinant estimator. *Technometrics* 41(3), 212 (1999)
11. Segata, N., Blanzieri, E.: Fast and scalable local kernel machines. *J. Mach. Learn. Res.* 11, 1883 (2009)
12. Woolrich, M.: Robust group analysis using outlier inference. *Neuroimage* 41, 286 (2008)

Segmenting Images by Combining Selected Atlases on Manifold

Yihui Cao¹, Yuan Yuan¹, Xuelong Li¹, Baris Turkbey², Peter L. Choyke²,
and Pingkun Yan¹

¹ Center for OPTical IMagery Analysis and Learning (OPTIMAL), State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, Shaanxi, P.R. China

² National Institutes of Health, National Cancer Institute, Molecular Imaging Program, Bethesda, MD 20892, USA

Abstract. Atlas selection and combination are two critical factors affecting the performance of atlas-based segmentation methods. In the existing works, those tasks are completed in the original image space. However, the intrinsic similarity between the images may not be accurately reflected by the Euclidean distance in this high-dimensional space. Thus, the selected atlases may be away from the input image and the generated template by combining those atlases for segmentation can be misleading. In this paper, we propose to select and combine atlases by projecting the images onto a low-dimensional manifold. With this approach, atlases can be selected according to their intrinsic similarity to the patient image. A novel method is also proposed to compute the weights for more efficiently combining the selected atlases to achieve better segmentation performance. The experimental results demonstrated that our proposed method is robust and accurate, especially when the number of training samples becomes large.

1 Introduction

Radiation therapy is often used in the treatment of cancers. It is important to acquire the accurate location of the target organ during the therapy. In clinical routine, this localization task is often performed by manually segmenting a series of images of a patient. However, manual segmentation is a tedious and time consuming procedure. In recent year, atlas-based methods, for their full automation and high accuracy, have become widely used approaches for medical image segmentation [1].

Generally, an atlas consists of a raw image and a corresponding label image. The basic idea of atlas based segmentation is that if a raw image of atlas is highly similar to the target image, the corresponding label image can be used to segment the target image through mapping. However, in practice, it is difficult to find a highly similar atlas. In the existing works, each atlas is first matched to the target image, resulting in a deformed image close to the target image. Then the “most similar” individual atlas is selected and used for segmentation [1]. It is

also straightforward to extend the scheme to multiple atlases, where more than one atlases can be selected and combined for segmentation [2].

It has been shown that using multiple atlases can yield more accurate results than using a single atlas [3]. Aljabar et al. [4] investigated different atlas selection strategies and showed that selection strategy is one of the significant factors affecting the performance. Roche et al. [5] studied three intensity-based image similarity measurements (mean square distance, correlation coefficient, mutual information) and showed that mutual information is the “best” option assuming a statistical relationship. As for multiple atlases combination, the Weighted Voting algorithm was the most widely used method in the previous works [2,3], where computing the optimal weights of the corresponding atlases is the key in the such an algorithm.

Existing works mainly addressed the problems of atlas selection and weighted combination in the high-dimensional image space [1,2,3]. However, it was shown that the Euclidean distance in high-dimensional space may not accurately reflect the intrinsic similarity between the images [6]. In addition, it showed that the geodesic distance in the high-dimensional manifold space is a better measurement for computing the similarity between images.

However, in practice, it is difficult to compute the geodesic distance directly in the high-dimensional space. An approach to solve this problem is to project the high-dimensional data onto a low-dimensional manifold space and preserving the local geometry in the same time using nonlinear dimensionality reduction techniques [6,7,8]. In such a low-dimensional space, the Euclidean distance can approximately reflect the intrinsic similarity between the images. Based on this fact, in this paper, we propose a new method to select atlases by relying on the intrinsic similarity between the images. In addition, we develop a novel method to compute the weights for combining the selected images into a single template to achieve better segmentation performance. The basic idea is to reconstruct the input patient image using the selected images in the low-dimensional manifold space.

The rest of this paper is organized as follows. In Section 2, we describe the workflow of our method and present the proposed atlas selection and combination methods. Our data and experimental results are provided in Section 3. Some conclusions about our work are drawn in Section 4.

2 Method

In this section, we briefly show the workflow of our proposed method as shown in Fig. 1. The upper row of the figure shows the process about the raw images of atlas, and the lower row shows the process about the label images of atlas. In our work, the raw image of atlas is denoted by A and the corresponding label image is denoted by L . The image to be segmented namely the patient image is denoted by P . The whole method includes three main steps: atlas registration, atlas selection, and atlas combination.

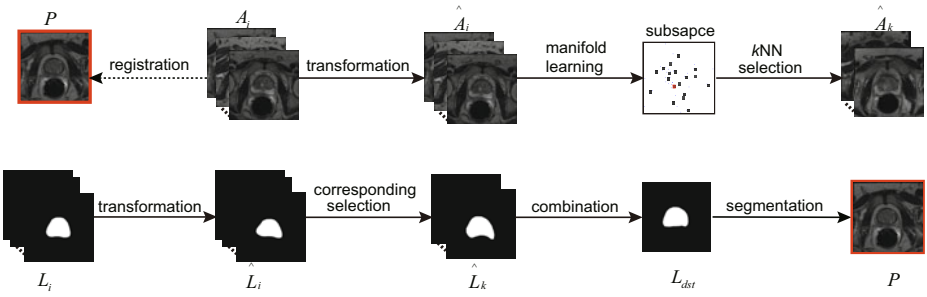


Fig. 1. The workflow of our method. The upper row shows the process for analyzing the raw images of the atlases, and the lower row shows the processing steps of the corresponding label images.

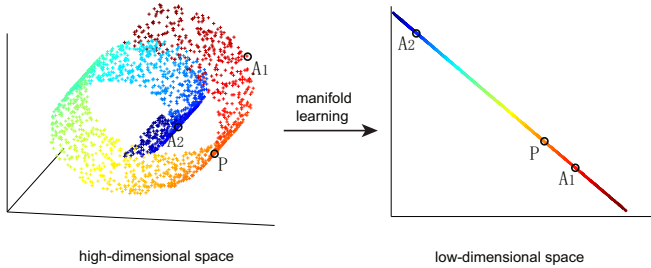


Fig. 2. An example illustrating the motivation of using the manifold learning for atlas selection in our method. In the Euclidean space the distance $PA_1 > PA_2$, while in manifold space $PA_1 < PA_2$.

2.1 Atlas Registration

The first stage is registration that each raw image of atlas A_i ($i = 1, 2, \dots, N$) is matched to the patient image P . The registration includes a 2D rigid registration and a non-rigid B-spline deformable registration. It yields a set of transformation parameters during the registration. Using the transformation parameters, each raw image A_i and its corresponding label image L_i are transformed to \hat{A}_i and \hat{L}_i , respectively, which are close to the patient image.

2.2 Manifold Learning for Atlas Selection

In terms of the theory of manifold learning, the raw images of atlas and the patient image can form a manifold in high-dimensional space, namely image space [6, 7, 8]. Take the “Swiss roll” structure as an example to illustrate the motivation of using manifold learning for atlas selection, as shown in Fig. 2. The points in the left 3D “Swiss roll” represent the images in the high-dimensional manifold. And the points (P, A_1, A_2) in the manifold respectively represents a

patient image and two atlases. It can be seen that in the Euclidean space the distance $PA_1 > PA_2$, while in manifold space $PA_1 < PA_2$. By manifold learning techniques, the high-dimensional data can be projected onto a low-dimensional space preserving the neighborhood information as shown in the right of graph. It can be seen that the distances of point P to the other points can be apparently reflected in the low-dimensional space. That is to say, the intrinsic similarity of images is better reflected in low-dimensional space than in high-dimensional space.

There were many classical manifold learning algorithms, such as the linearly projection of Principal Component Analysis (PCA) and the nonlinear ISOMAP [6] and LLE [7] algorithms. In this work, we applied the Locality Preserving Projections (LPP) [8] algorithm in manifold learning. For it share the advantages of both linear and nonlinear algorithms, namely linearly projecting high-dimensional data onto a low-dimensional space and preserving local neighborhood information as well.

In our work, we set all images to the same size of m pixels. And each raw image of atlas is represented by a high-dimensional vector \mathbf{h}_i the size of which is m . Then the high m -dimensional vectors \mathbf{h}_i are projected to low n -dimensional vectors \mathbf{l}_i ($n \ll m$). The projection yields a transformation matrix \mathbf{W} . Then the patient image P projects onto the same low-dimensional space. The projection is defined as follows:

$$\mathbf{p}_l = \mathbf{W}^T \times \mathbf{p}_h, \quad (1)$$

where the vectors \mathbf{p}_h and \mathbf{p}_l represent the patient image in high- and low-dimensional space, respectively.

In the low-dimensional space, K nearest vectors \mathbf{l}_k ($k = 1, 2, \dots, K$) of \mathbf{p}_l are selected by the Euclidean distance. Correspondingly, K most similar raw images of atlas \hat{A}_k to the patient image P are selected. Corresponding to \hat{A}_k , the K label images \hat{L}_k are selected from \hat{L}_i . And these selected label images will be used for combination in the next stage.

2.3 Atlas Combination

In atlas combination stage, we employ the robust Weighted Voting algorithm [2,3]. The algorithm is defined as follows:

$$L = \sum_{k=1}^K w_k \hat{L}_k, \quad (2)$$

where L is the result of combination image, and $\mathbf{w}=(w_1, w_2, \dots, w_K)$ is a vector representing the weights of the selected label images \hat{L}_k . Apparently the weights are the key in such the algorithm.

It assumes that the vectors \mathbf{l}_k are distributed in a linear low-dimensional space [7]. Combining the selected label images into a single template can be considered as using the vectors \mathbf{l}_k to reconstruct the vector \mathbf{p}_l . Thus, computing the optimal weights of combination can be solved by minimizing the linear reconstruction error [9]. The error ε is defined as follows:

$$\arg \min_{w_1, \dots, w_K} \varepsilon = \|\mathbf{p}_l - \sum_{k=1}^K w_k \mathbf{l}_k\|_2^2, \quad (3)$$

with the weights constraint $\sum w_k = 1$. Apparently, it is a constrained least squares problem. We use the Lagrange multiplier algorithm and introducing a Gram matrix \mathbf{G} to solve this problem.

$$\mathbf{G} = (\mathbf{p}_l \mathbf{1}^T - \mathbf{L})^T (\mathbf{p}_l \mathbf{1}^T - \mathbf{L}), \quad (4)$$

where $\mathbf{1}$ is a vector of ones with the same size of \mathbf{p}_l , \mathbf{L} is a matrix that consists of \mathbf{l}_k . Then the problem can be solved by the following solution:

$$\mathbf{w} = \frac{\mathbf{G}^{-1} \mathbf{1}}{\mathbf{1}^T \mathbf{G}^{-1} \mathbf{1}}. \quad (5)$$

Applying the weights \mathbf{w} to the Weighted Voting algorithm (2), it yields a single combination image L . Since the edge of L may not be smooth, several basic morphological methods are used to address the problem. Finally, it yields the template L_{dst} which is used for the final patient image segmentation.

3 Experiments

3.1 Data and Measurement

In our experiments, the proposed method was tested on 40 MR prostate images, which were taken from 40 different patients. Each image has 512×512 pixels. The binary label images, which were manually segmented by an expert, were considered as the ground truth. The Dice Similarity Coefficient (DSC) was used to evaluate the segmentation performance of our proposed method quantitatively. The evaluation criterion of DSC is defined as follows:

$$DSC(A, B) = \frac{2|A \cap B|}{|A| + |B|}, \quad (6)$$

where A is the ground truth image, B is the result of an automatically segmented image, and $|\cdot|$ denotes the area of the target region. The DSC value varies between 0 and 1 and a higher value indicates a better segmentation.

3.2 Results

Note that the atlas selection and the weights computation were all performed in a low-dimensional space in our method. In order to decrease the computation time, we extracted the region of interest around the image center with the size of 256×256 pixels. That was to say, all atlas images formed a 65536-dimensional manifold space. All data were projected onto a low-dimensional space using manifold learning. We evaluated the segmentation performance over dimensionality varying from 1 to 39, where 39 is the upper bound set by the number of training images. We found that the value of 38 yielded the best performance in our

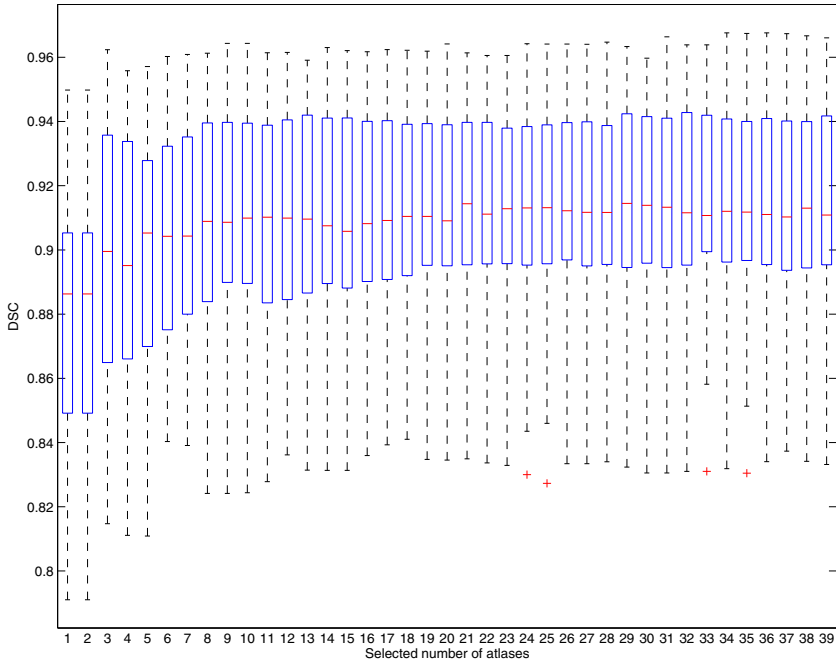


Fig. 3. A box-and-whisker diagram showing the distributions of our results

experiments, which was then adopted. Before the projection, all images were preprocessed by histogram equalization to reduce the influence of light. For the relatively limited data, a leave-one-out approach was employed to validate the performance of our method. At each time, one of the 40 images was used as the patient image and the rest were applied as atlases.

In our experiments, the selected number of atlas was the only variable parameter. Therefore, we gave the results based on the selected number of atlas changing from 1 to 39. In Fig. 3, it was a box-and-whisker diagram showing the distributions of our results. At each column, it depicts five-number summaries: the smallest observation, lower quartile, median, upper quartile, and largest observation. It can be seen that the median DSC varying between 0.88 and 0.92, and the values were larger than 0.90 in most cases.

We compared the performance of our method and the state-of-the-art method proposed by Klein et al. [2]. The main difference between the two methods is that the atlas selection and the weights computation in our method are all performed in a low-dimensional space. For a fair comparison, two methods were carried out based on the same registration. When the atlas selected number ranged from 1 to 39, the average DSC values of 40 experiments were shown in Fig. 4. It can be seen that the DSC values of our method were higher than Klein's in most cases. Especially, the problem of over-fitting apparently occurred in Klein's method

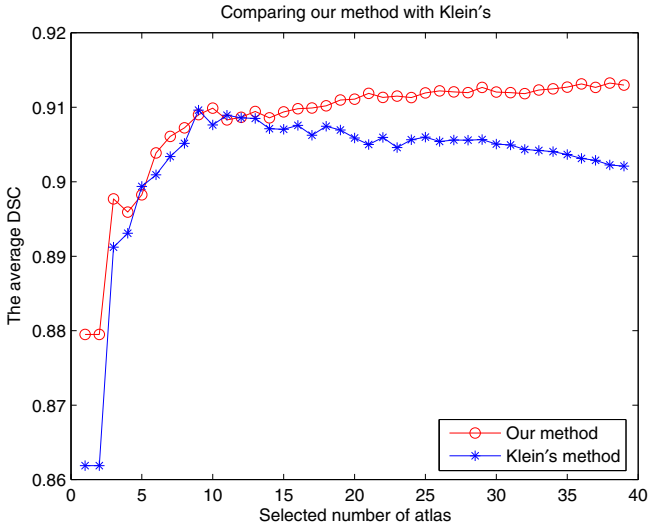


Fig. 4. The comparison of our method and Klein’s method by varying the number of atlases

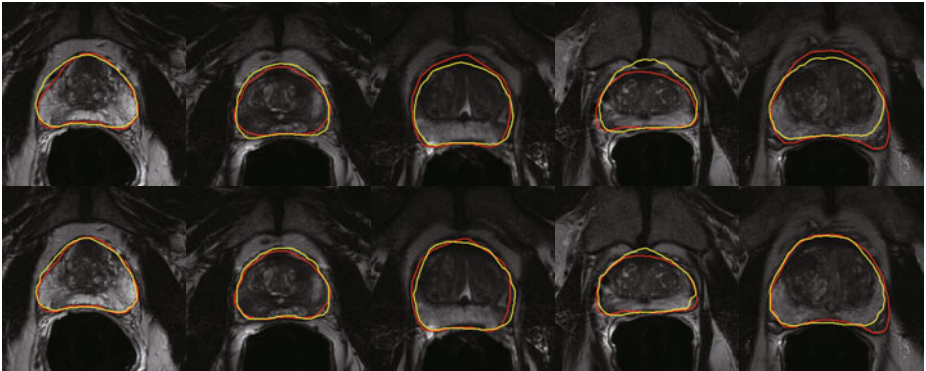


Fig. 5. Qualitative results of Klein’s (in top row) and our method (in bottom row). The red contours represent the ground truth. The yellow contours are automatically delineated by the two methods.

when the number of the samples exceeded a limitation. It has been shown that our method is superior to Klein’s method in the case of relative large sample ($p < 0.01$) and is not inferior in the case of small sample.

When we set the selected number of atlas to 35, several visual segmentation results were obtained and shown in Fig. 5. The qualitative results of Klein’s are in top row and our method’s are in bottom row. The red contours represent the ground truth, and the yellow contours are automatically delineated by the two methods.

4 Conclusion

In this paper, we proposed a novel atlas-based method for automatic medical image segmentation. We proposed to select atlases according to the intrinsic similarity to the patient image in the low-dimensional space. A novel method was also proposed to compute the weights for more efficiently combining the selected atlases to achieve better segmentation performance. By comparing with the state-of-the-art method [2], the experimental results show our method is robust and promising. In future work, we will test our algorithm on other datasets and further extend our method to 3D medical images segmentation.

Acknowledgment. The presented research work is supported by the National Basic Research Program of China (973 Program) (Grant No. 2011CB707000) and the National Natural Science Foundation of China (Grant No. 61072093).

References

1. Wu, M., Rosano, C., Lopez-Garcia, P., Carter, C.S., Aizenstein, H.J.: Optimum template selection for atlas-based segmentation. *NeuroImage* 34(4), 1612–1618 (2007)
2. Klein, S., van der Heide, U.A., Lips, I.M., van Vulpen, M., Staring, M., Pluim, J.P.W.: Automatic segmentation of the prostate in 3D MR images by atlas matching using localized mutual information. *Med. Phys.* 35(4), 1407–1417 (2008)
3. Artaechevarria, X., Muñoz-Barrutia, A., Ortiz-de-Solórzano, C.: Combination strategies in multi-atlas image segmentation: Application to brain MR data. *IEEE Trans. Medical Imaging* 28(8), 1266–1277 (2009)
4. Aljabar, P., Heckemann, R., Hammers, A., Hajnal, J., Rueckert, D.: Classifier selection strategies for label fusion using large atlas databases. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) *MICCAI 2007, Part I. LNCS*, vol. 4791, pp. 523–531. Springer, Heidelberg (2007)
5. Roche, A., Malandain, G., Ayache, N.: Unifying maximum likelihood approaches in medical image registration. *Int. J. Imag. Syst. Technol.*, 71–80 (2000)
6. Tenenbaum, J.B., de Silva, V., Langford, J.C.: A global geometric framework for nonlinear dimensionality reduction. *Science* 290 (2000)
7. Roweis, S., Saul, L.K.: Nonlinear dimensionality reduction by locally linear embedding. *Science* 290 (2000)
8. He, X., Niyogi, P.: Locality preserving projections. In: *Proc. Neural Information Processing Systems* (2003)
9. Chang, H., Yeung, D.Y., Xiong, Y.: Super-resolution through neighbor embedding. In: *Proc. IEEE CVPR*, vol. 1, pp. 275–282 (2004)

Learning Pit Pattern Concepts for Gastroenterological Training

Roland Kwitt^{1,*}, Nikhil Rasiwasia², Nuno Vasconcelos², Andreas Uhl¹,
Michael Häfner³, and Friedrich Wrba⁴

¹ Dept. of Computer Sciences, Univ. of Salzburg, Austria

² Dept. of Electrical and Computer Engineering, Univ. of California, San Diego, USA

³ Dept. of Internal Medicine, St. Elisabeth Hospital, Vienna, Austria

⁴ Dept. of Clinical Pathology, Vienna Medical Univ., Austria

Abstract. In this article, we propose an approach to learn the characteristics of colonic mucosal surface structures, the so called *pit patterns*, commonly observed during high-magnification colonoscopy. Since the discrimination of the pit pattern types usually requires an experienced physician, an interesting question is whether we can automatically find a collection of images which most typically show a particular pit pattern characteristic. This is of considerable practical interest, since it is imperative for gastroenterological training to have a representative image set for the textbook descriptions of the pit patterns. Our approach exploits recent research on semantic image retrieval and annotation. This facilitates to learn a semantic space for the pit pattern concepts which eventually leads to a very natural formulation of our task.

1 Motivation

Over the past few years there has been considerable research in computer-based systems to guide *in vivo* assessment of colorectal polyps, using endoscopic imaging. This research is motivated by the prevalence of colorectal cancer, one of the three most commonly diagnosed forms of cancer in the US, and its high mortality rate. Following the concept of the *adenoma-carcinoma sequence* [11], colorectal cancer predominantly develops from adenomatous polyps, although adenomas do not inevitably become cancerous. In fact, the resection of colorectal adenomas reduces the incidence of colorectal cancer. In this context, it is safe to say that the ultimate objective of image analysis is to distinguish *neoplastic* from *non-neoplastic* lesions, although finer grained discriminations are obviously possible. While early approaches (e.g. [5]) to computer-assisted dignity assessment were based on visual data from conventional white-light endoscopes, research has shifted towards novel imaging modalities. These include narrow band imaging (NBI, e.g. [9]), high-magnification chromo-endoscopy (HMCE e.g. [4]), and probe-based confocal laser endomicroscopy (e.g. [1]). The emergence of these

* This work is partially funded by the Austrian Science Fund (FWF) under Project No. L366-N15.

novel imaging modalities has made it challenging for gastroenterologists to interpret the acquired imagery. In order to prevent serious mistakes (e.g. perforation of the colon, etc.), substantial experience with the particular imaging modality and the highlighted tissue structures is still necessary, especially in situations where the physician’s assessment differs from that of the decision support system (which is still an uncommon tool in clinical practice).

In this work, we tackle the problem of preparing prospective gastroenterologists for clinical practice with the novel imaging modalities. We argue that, during gastroenterological training, it is imperative to have (i) access to a database of labeled images from the prospective imaging modality and (ii) possibility to browse through images depicting the *textbook description* of a particular structure. In the absence of a computer vision system to assemble these images, an experienced gastroenterologist will typically have to work through a vast image repository, to sort out the most relevant training examples. We propose a computer vision solution to this problem, based on recent advances in semantic image retrieval [8]. This is a formulation of image database search, where images are mapped onto a *semantic space* of image *concepts*.

While, in computer vision, concepts are usually cars or buildings, the idea can be applied to the pit pattern classes commonly used in the medical literature for prediction of histopathological results (cf. [3]). Unlike [8], we are not interested in the strict task of retrieval by semantic example. Instead, our work is directed to the semantic browsing scenario. This is the scenario where gastroenterologists are able to browse the image space efficiently by focusing on regions where particular concepts, i.e. pit patterns, are most prominent. We propose a system that enables this type of *semantically focused browsing*. Although our approach is generic, we demonstrate its applicability in the context of HMCE and Kudo’s pit pattern analysis scheme [6].

The technical details of pit pattern browsing are given in the following section. Section 3 is devoted to experimental results and Sect. 4 presents our conclusions.

2 Learning the Pit Pattern Concepts

The starting point of our approach is a database of endoscopy images $\mathcal{D} = \{I_1, \dots, I_{|\mathcal{D}|}\}$ and a collection of concepts $\{w_1, \dots, w_C\}$, i.e. the pit pattern types. We require that each database image is augmented by a binary *caption* vector $\mathbf{c}_y \in \{0, 1\}^C$, where $c_y^j = 1$ signifies that the j -th concept is present in image I_y . This is termed a *weakly* labeled set of images, since $c_y^j = 0$ does not necessarily mean that the j -th concept is not present. We further don’t know which image region contains the annotated concept (i.e. no prior segmentation available). In fact, weak labeling is carried to the extreme, since the caption vectors only contain one non-zero entry for the prominent concept. This follows from the fact that the medical labeling procedure is based on reconciliation with histopathological ground-truth. For example, if the laboratory results indicate a normal gland and the gastroenterologist has the visual impression of a pit pattern type I, then the image is labeled with that type. However, a pit pattern of type II

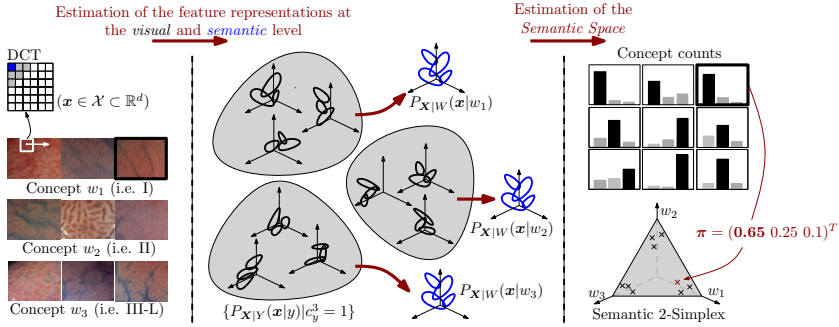


Fig. 1. Processing pipeline for learning three exemplary pit pattern concepts $\{w_1, w_2, w_3\}$ (e.g. I, II, III-L). First, we decompose each image into a collection of *localized* features. Then, we estimate (i) the *semantic-level* feature representations $\{P_{\mathbf{X}|W}(\mathbf{x}|w_i)\}_{i=1}^3$ from the *visual-level* feature representations and (ii) the mapping from feature space to *semantic space*, i.e. the semantic 2-simplex embedded in \mathbb{R}^3 .

(or any other type) can also be visible in some areas of the image. While this labeling strategy guarantees that the annotated pit pattern is visible to some extent, at least to the experienced gastroenterologist, many of the images do not convey the textbook description [6] corresponding to the labeled pit pattern.

2.1 Image Representation at the Visual and Semantic Level

The first stage for learning the image concepts is similar to previous studies (e.g. [7]), where automated dignity assessment was the primary objective. Each image I in the database is represented by a collection of *localized* features $I = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ drawn independently from a random vector \mathbf{X} , defined in some feature space $\mathcal{X} \subset \mathbb{R}^d$. This stage is illustrated in the leftmost part of Fig. 1. Defining a random variable Y (with realizations in $\{1, \dots, |D|\}$) such that $Y = y$ when features are drawn from image I_y , the probability of image I at the *visual* level is

$$P_{\mathbf{X}|Y}(I|y) = \prod_{j=1}^N P_{\mathbf{X}|Y}(\mathbf{x}_j|y) . \tag{1}$$

The density (i.e. the generative model) $P_{\mathbf{X}|Y}(\mathbf{x}|y)$ for image I_y is estimated by a K_V -component multivariate Gaussian mixture

$$P_{\mathbf{X}|Y}(\mathbf{x}|y) = \sum_{k=1}^{K_V} \gamma_y^k \mathcal{G}(\mathbf{x}; \boldsymbol{\mu}_y^k, \boldsymbol{\Sigma}_y^k) \text{ with } \sum_k \gamma_y^k = 1 , \tag{2}$$

based on the corresponding collection of features.

In contrast to [7], where captions are neglected, and images retrieved by visual similarity (query-by-visual-example), the captions now represent a key component of the system. Introducing a random variable W (with realizations in

$\{1, \dots, C\}$) such that $W = i$ when features are drawn from concept w_i , induces a new collection of probability densities $\{P_{\mathbf{X}|W}(\mathbf{x}|w_i)\}_{i=1}^C$ on \mathcal{X} . These densities are denoted the feature representations at *semantic* level. Assuming conditional independence of the features given concept membership, the *concept-conditional* probability of image I at the *semantic* level is

$$P_{\mathbf{X}|W}(I|w) = \prod_{j=1}^N P_{\mathbf{X}|W}(\mathbf{x}_j|w) . \tag{3}$$

Similar to density estimation at the visual level, we use multivariate Gaussian mixtures with K_S components to estimate $P_{\mathbf{X}|W}(\mathbf{x}|w)$, i.e.

$$P_{\mathbf{X}|W}(\mathbf{x}|w) = \sum_{l=1}^{K_S} \alpha_w^l \mathcal{G}(\mathbf{x}; \boldsymbol{\nu}_w^l, \boldsymbol{\Phi}_w^l) \text{ with } \sum_l \alpha_w^l = 1 . \tag{4}$$

Modeling the densities at the visual *and* semantic level by Gaussian mixtures has the convenient advantage that we can exploit the hierarchical mixture modeling approach of [10] to estimate the mixture parameters at the semantic level $\{\alpha_w^l, \boldsymbol{\nu}_w^l, \boldsymbol{\Phi}_w^l\}$ from the mixture parameters at the visual level $\{\lambda_y^k, \boldsymbol{\mu}_y^k, \boldsymbol{\Sigma}_y^k\}$. This step is visualized in the middle of Fig. 1, where the mixtures associated with several images in a class are summarized by the class' single semantic-level mixture. Note that the number of Gaussian components at semantic level ($C \times K_S$) is considerably smaller than the number of Gaussian components at visual level ($|\mathcal{D}| \times K_V$). Figure 1 illustrates the case where $K_V = 3$ and $K_S = 4$. The computational effort to estimate the semantic-level mixtures, using the method of [10], is also considerably smaller than that required for direct estimation of $P_{\mathbf{X}|W}(\mathbf{x}|w)$ based on the pooled features of all images annotated with concept w .

2.2 Learning the Semantic Space

The identification of images which most characteristically depict a particular concept requires a semantic image representation with explicit control over the concepts. In [8], the authors introduce the idea of an image as a point on a *semantic space*. The image is first modeled as a vector $I_y = (n_y^1, \dots, n_y^C)^T$ of concept counts (cf. top right part of Fig. 1), where n_y^k is the number of feature vectors in the y -th image drawn from the k -th concept. The concept count vectors are then modeled as realizations of a multinomial random variable \mathbf{T} . As illustrated on the bottom right of Fig. 1, the parameter vector $\boldsymbol{\pi}_y = (\pi_y^1, \dots, \pi_y^C)^T$ of the multinomial distribution associated with the image is a point on the standard $(C - 1)$ -simplex, since $\sum_i \pi^i = 1$. This simplex is denoted the *semantic space*, and $\boldsymbol{\pi}_y$ the *semantic multinomial (SMN)* associated with image I_y .

The question is how to estimate the mapping based on a database of tuples $\{(I_y, \mathbf{c}_y)\}_{y=1, \dots, |\mathcal{D}|}$. For that purpose, we employ a modification of the semantic multiclass labeling approach of [2] which implements the mapping based on an estimation of posterior concept probabilities, i.e.

$$\pi_y^w = P_{W|\mathbf{X}}(w|I_y) = \frac{P_{\mathbf{X}|W}(I_y|w)P_W(w)}{P_{\mathbf{X}}(I_y)} . \tag{5}$$

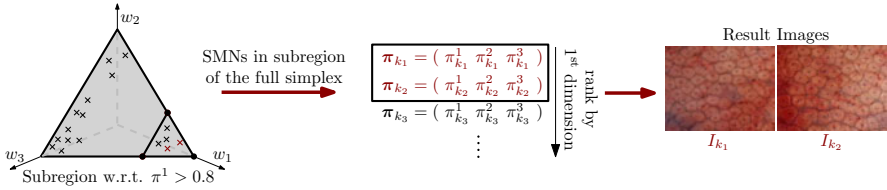


Fig. 2. Identifying the images, represented by SMNs, which most typically represent the concept w_1 (here pit pattern type I)

Although it is possible to directly estimate π_y^w by assuming a uniform concept prior $P_W(w)$ and estimating $P_X(I_y)$ by $\sum_w P_{X|W}(I_y|w)P_W(w)$, we choose an alternative approach to cover for numerical instabilities. The strategy is to compute $\log P_{W|X}(w|\mathbf{x}_j)$ for each feature vector \mathbf{x}_j separately, using (4), then determine the concept with the largest posterior probability per feature vector and eventually tally the occurrences of the *winning* concepts. This facilitates Maximum-Likelihood (ML) parameter estimation of π_y through $\forall w : \hat{\pi}_y^w = n_y^w \cdot 1/N$. In [8], the authors further suggest regularization with a Dirichlet prior, which leads to a maximum-a-posteriori estimate

$$\hat{\pi}_y^w = \frac{\tilde{\pi}_y^w + \pi_0}{\sum_{i=1}^C (\tilde{\pi}_y^i + \pi_0)}, \tag{6}$$

where π_0 is a regularization parameter. The remaining question is how to identify the desired set of images from the semantic representation in the form of points on the semantic space. Given that we aim to identify the images most typical of the concept w_i (e.g. pit pattern III-L), we only need to navigate on the simplex. In fact, we can easily identify a subregion of the full simplex whose SMNs represent images where the w_i -th concept is prominent with probability t , by using $\pi^i > t, t \in [0, 1]$. Fig. 2 illustrates this idea for $\pi^1 > 0.8$. Sorting the SMNs in that region along the i -th dimension gives a list of the most representative (i.e. top-ranked) images for concept w_i .

3 Experimental Evaluation

We evaluate our proposed approach on a weakly labeled database of 716 HMCE images (magnified $150\times$) of size 256×256 pixel, captured by an Olympus Evis Exera CF-Q160ZI/L endoscope. The images stem from a total of 40 patients and the database contains only images where the histological ground truth is coherent with the annotated pit pattern type. The cardinalities of the image sets per concept are $\{124, 74, 124, 20, 276, 98\}$ for pit pattern types I, II, III-L, III-S, IV and V. A graphical and textual description of the pit pattern characteristics is provided in Fig. 3. Apart from the gastroenterologist’s experience, these descriptions represent the *textbook* material for dignity assessment of colorectal lesions. The images are converted from RGB to YBR color space for further processing.

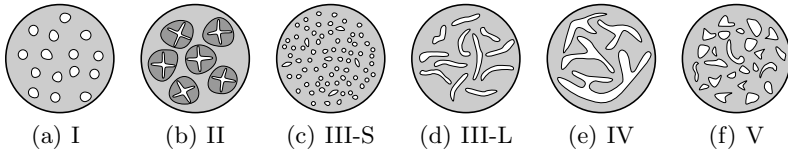


Fig. 3. Schematic illustration of the six pit pattern types according to [6]. The typical characteristics are (a) normal, round pit; (b) asteroid, stellar or papillary pit; (c) tubular or round pit (smaller than type I); (d) tubular or round pit (larger than type I); (e) dendritic or gyrus-like pit; (f) irregular arrangements, loss of structure.

As *localized* features, we use DCT coefficients (extracted in zigzag scan order) of 8×8 patches, obtained from a sliding window, moving by two pixel increments in both image dimensions (cf. Fig. 1). We extract the first 16 coefficients (including the DC coefficient) from the same patch across the color channels and arrange the coefficients in feature vectors according to a YBRYBRYBR... interleaving pattern. The Gaussian mixtures to model $P_{\mathbf{X}|Y}(\mathbf{x}|y)$ are fitted by the classic Expectation-Maximization (EM) algorithm. The number of mixture components at this level is set to $K_V = 8$ and we restrict the covariance matrices to diagonal form. At the semantic level, we set $K_S = 64$ and estimate the parameters using the hierarchical estimation approach of [10]. Regarding SMN estimation, we choose a regularization parameter $\pi_0 = 1/6$, although experiments show that the approach is not sensitive to this choice.

To demonstrate that we can actually identify images which most typically depict the textbook pit pattern descriptions, we sort the SMNs on the semantic simplex along the dimension corresponding to each concept and extract the K top-ranked images. We further ensure that the extracted images do *not* belong to the same patient in order to establish a realistic scenario. We refer to this step as *patient pruning* of the result set. Figure 4 shows the images after pruning the $K = 10$ top-ranked images per pit pattern concept. Due to the fact that the database images are not uniformly distributed over the patients, the pruning step has the effect that the cardinality of the final browsing result per concept is not equal. Nevertheless, a comparison to the illustrations and descriptions in Fig. 3 reveals the correspondences we were looking for: we observe the characteristic gyrus-like structures of pit pattern IV, the round pits of pit pattern I, or the complete loss of structure in case of pit pattern V for instance.

Besides visual inspection of the results in Fig. 4 we conduct a more objective evaluation by exploiting the ground-truth caption vectors for each image. In particular, we evaluate the *average error rate* of the system when browsing the K top-ranked images per concept. We perform a *leave-one-patient-out (LOPO)* evaluation, where we adhere to the following protocol per patient: (i) remove the patient's images from the database, (ii) estimate the SMNs based on the remaining images, (iii) extract the K top-ranked images (now using the whole database) per concept and (iv) perform the patient pruning step. The average error rate is then calculated as the percentage of images (averaged over all LOPO

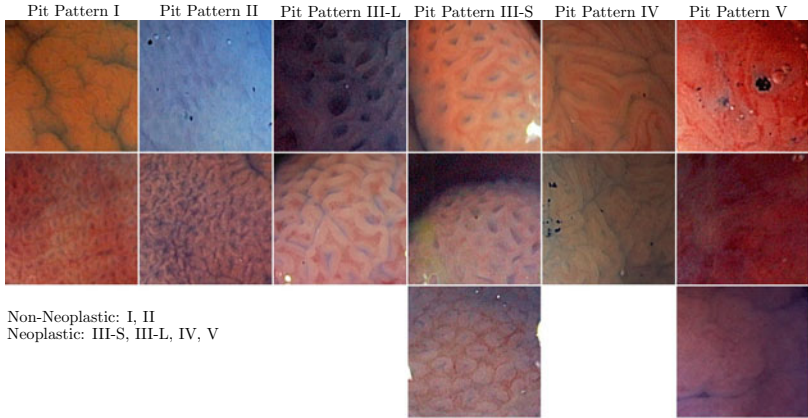


Fig. 4. Result of identifying the most representative images for each pit pattern concept at the operating point $K = 10$ (with patient pruning)

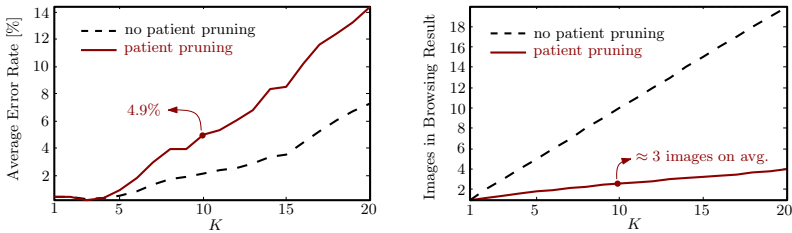


Fig. 5. Average error rate with respect to the ground-truth caption vectors, when browsing the K top-ranked images of each pit pattern type

runs) in the final browsing result of concept w_i which do not belong there according to the corresponding ground-truth caption vectors (i.e. zero entry at the i -th position). Figure 5 shows the average error rate in dependence of K with and without patient pruning (for comparative reasons). At the operating point $K = 10$ for instance, we obtain three images per concept on average at an error rate of 4.9%. This corresponds to ≈ 0.88 wrong images in the final browsing result.

4 Concluding Remarks

Motivated by the need to provide prospective gastroenterologists with a collection of images showing the most typical characteristic of a particular mucosal structure during endoscopy, we presented a generic approach to establish a semantic space on a database of weakly labeled HMCE images. To the best of our knowledge, introducing the notion of a semantic domain to that problem has not been done so far. On the basis of Kudo’s pit pattern analysis scheme,

we demonstrated that browsing the semantic space in *interesting* regions in fact allows to isolate the most characteristic images for each pit pattern type.

References

1. André, B., Vercauteren, T., Perchant, A., Buchner, A.M., Wallace, M.B., Ayache, N.: Endomicroscopic image retrieval and classification using invariant visual features. In: Proceedings of the IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI 2009), Boston, MA, USA, pp. 346–349 (June 2009)
2. Carneiro, G., Vasconcelos, N.: Formulating semantic image annotation as a supervised learning problem. In: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR 2005), San Diego, CA, USA, pp. 163–168 (June 2005)
3. East, J.E., Suzuki, N., Saunders, B.P.: Comparison of magnified pit pattern interpretation with narrow band imaging versus chromoendoscopy for diminutive colonic polyps: A pilot study. *Gastrointest. Endosc.* 66(2), 310–316 (2007)
4. Häfner, M., Gangl, A., Kwitt, R., Uhl, A., Vécsei, A., Wrba, F.: Improving pit-pattern classification of endoscopy images by a combination of experts. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5761, pp. 247–254. Springer, Heidelberg (2009)
5. Karkanis, S.A., Iakovidis, D., Karras, D., Maroulis, D.: Detection of lesions in endoscopic video using textural descriptors on wavelet domain supported by artificial neural network architectures. In: Proceedings of the IEEE International Conference on Image Processing (ICIP 2001), Thessaloniki, Greece, pp. 833–836 (October 2001)
6. Kudo, S., Hirota, S., Nakajima, T., Hosobe, S., Kusaka, H., Kobayashi, T., Himori, M., Yagyuu, A.: Colorectal tumours and pit pattern. *J. Clin. Pathol.* 47(10), 880–885 (1994)
7. Kwitt, R., Uhl, A., Häfner, M., Gangl, A., Wrba, F., Vécsei, A.: Predicting the histology of colorectal lesions in a probabilistic framework. In: Proceedings of the IEEE International Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA 2010), San Francisco, CA, USA, pp. 103–110 (June 2010)
8. Rasiwasia, N., Moreno, P., Vasconcelos, N.: Bridging the gap: Query by semantic example. *IEEE Trans. Multimedia* 9(5), 923–938 (2007)
9. Tischendorf, J.J.W., Gross, S., Winograd, R., Hecker, H., Auer, R., Behrens, A., Trautwein, C., Aach, T., Stehle, T.: Computer-aided classification of colorectal polyps based on vascular patterns: a pilot study. *Endoscopy* 42(3), 203–207 (2010)
10. Vasconcelos, N., Lippman, A.: Image indexing with mixture hierarchies. In: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR 2001), Kauai, HI, USA, pp. 3–10 (December 2001)
11. Vogelstein, B., Fearon, E.R., Hamilton, S.R., Kern, S.E., Preisinger, A.C., Leppert, M., Nakamura, Y., White, R., Smits, A.M., Bos, J.L.: Genetic alterations during colorectal-tumor development. *N. Engl. J. Med.* 319(9), 525–532 (1988)

Immediate Structured Visual Search for Medical Images

Karen Simonyan¹, Andrew Zisserman¹, and Antonio Criminisi²

¹ University of Oxford, UK

{karen,az}@robots.ox.ac.uk

² Microsoft Research, Cambridge, UK
antcrim@microsoft.com

Abstract. The objective of this work is a scalable, *real-time* visual search engine for medical images. In contrast to existing systems that retrieve images that are globally similar to a query image, we enable the user to select a query Region Of Interest (ROI) and automatically detect the corresponding regions within all returned images. This allows the returned images to be ranked on the content of the ROI, rather than the entire image. Our contribution is two-fold: (i) immediate retrieval – the data is appropriately pre-processed so that the search engine returns results in real-time for any query image and ROI; (ii) structured output – returning ROIs with a choice of ranking functions. The retrieval performance is assessed on a number of annotated queries for images from the IRMA X-ray dataset and compared to a baseline.

1 Introduction

The exponential growth of digital medical images of recent years poses both challenges and opportunities. Medical centres now need efficient tools for analysing the plethora of patient images. Myriads of archived scans represent a huge source of data which, if exploited, can inform and improve current clinical practice.

This paper presents a new, scalable, algorithm for the immediate retrieval of medical images *and* structures of interest within them: given a query image and a specified region of interest (ROI) we return images with the corresponding ROI (e.g. the same bone in the hand) delineated. The returned images can be ranked on the contents of the ROI.

Why Immediate Structured Image Search? Given a patient with a condition (e.g. a tumour in the spine) retrieving other generic spine X-rays may not be as useful as returning images of patients with the same pathology, or of exactly the same vertebra. The structured search with an ROI is where we differ from existing content-based medical image retrieval methods which return images that are *globally* similar to a query image [10]. The immediate aspect of our work enables a flexible exploration as it is not necessary to specify in advance what region (e.g. an organ or anomaly), to search for – every region is searchable.

Use cases include: conducting population studies on specific anatomical structures; tracking the evolution of anomalies efficiently; and finding similar anomalies or pathologies in a particular region. The ranking function can be modified to order the returned images according to the similarity between the query and target ROI's shape or image content. Alternatively, the ROI can be classified, e.g. on whether it contains a particular anomaly such as cysts on the kidney, or arthritis in bones, and ranked by the classification score.

Outline. Sect. 2 describes the retrieval algorithm using X-rays of hands as the running example, with section Sect. 2.4 giving examples of ROI ranking functions. Sect. 3 assesses the retrieval performance and compares to a baseline using images from the publicly available IRMA dataset [3], and Sect. 4 concludes.

Related work and challenges. The problem of ROI retrieval in medical images is addressed in [8], but only in the limiting case of manually pre-segmented ROIs. Our approach is inspired by the Video Google work of [12] for object localisation in videos, and later developments of ROI-driven image retrieval in computer vision for search in large scale image datasets [11]. However, the direct application of these techniques to medical images is not feasible (as shown in Sect. 3) because the feature matching and registration methods of these previous works do not account for inter-subject non-rigid transformations and the repeating structures common to medical images (e.g. phalanx or spine bones). Instead, we employ a registration method tuned to medical images, related to [1]. We differ in that we utilise feature point (landmark) descriptors based on the intensity derivatives, with matching guided to robustly fit non-rigid Thin Plate Spline (TPS) transformations.

2 Structured Image Retrieval Framework

The key to our approach is that registrations are pre-computed (off-line) so that at run time correspondences in target images can be determined immediately for any ROI in the query image. In the case of medical images it is possible to compute registrations between images if they are of the same class, e.g. if they are both images of hands. Given query image and ROI, three stages are involved: (i) image classification, so that ROI correspondences are only considered between images of the same class; (ii) approximate global registration for images within that class, this is pre-computed; and (iii) refinement of the ROI in a target image using the approximate registration as a guide. This is performed at run time. Fig. 1 summarizes the off-line and on-line parts of the framework.

We describe in the following sections how the images are classified; how the global registration is performed; and how the ROI in the target image is refined. Once the ROIs have been accurately localized in each image of the target set, there is then a choice of how these images should be ranked. We describe a number of possibilities for ranking functions in Sect. 2.4.

In the case of a dataset where new images are added, it is important that the method is scalable so that adding new images and using them to make a query

1. **On-line (given a user-specified query image and ROI bounding box)**
 - Select the target image set (repository images of the same class as the query).
 - Using the pre-computed registration (Sect. 2.2), compute the ROIs corresponding to the query ROI in all images of the target set.
 - Refine the ROIs using local search (Sect. 2.2).
 - Rank the ROIs using the similarity measure of choice (Sect. 2.4).
2. **Off-line (pre-processing)**
 - Classify the repository images into a set of pre-defined classes (Sect. 2.1).
 - Compute the registration for all pairs of images of the same class (Sect. 2.2).

Fig. 1. The off-line and on-line parts of the structured retrieval algorithm

does not cause a delay. We have achieved this by registering the query image to only a subset of the same class images; the transformation to the rest can be readily obtained by transform composition as described in Sect. 2.3.

2.1 Image Classification

Our aim is to divide the X-ray images into five classes: **hand**, **spine**, **chest**, **cranium**, **negative (the rest)**. Certain image retrieval methods take the textual image annotation into account. However, as shown in [5], the error rate of DICOM header information is high, which makes it infeasible to rely on text annotation for classification. Therefore, we use automated visual classification.

We employ the multiple kernel (MKL) technique of [13] and train a set of binary SVM classifiers on multi-scale dense-SIFT and self-similarity visual features in the “one-vs-rest” manner. The classified image is assigned to the class whose classifier outputs the largest decision value. The MKL formulation can exploit different, complementary image representations, leading to high-accuracy classification. The classifier is learnt from training images, and its accuracy is evaluated on a ground truth data set as described in Sect. 3.

2.2 Efficient and Robust Image Registration

In this section we first describe the registration algorithm for a pair of images. This algorithm is the basic workhorse that is used to compute registrations between all images (of the same class). We postpone until Sect. 2.3 how this is done in an efficient and scalable manner. In our case the registration method should be robust to a number of intraclass variabilities of our dataset (e.g. child vs adult hands) as well as additions and deletions (such as overlaid writing, or the wrists not being included). At the same time, it should be reasonably efficient to allow for the fast addition of a new image to the dataset. The method adopted here is a sequence of robust estimations based on sparse feature point matching. The process is initialized by a coarse registration based on matching the first and second order moments of the detected feature points distribution.

This step is feasible since the pairs of images to be registered belong to the same class and similar patterns of detected points can be expected. Given this initial transform T_0 , the algorithm then alternates between feature matching (guided by the current transform) and Thin-Plate Spline (TPS) transform estimation (using the current feature matches). These two stages are described next. We use single-scale Harris feature points, and the neighbourhood of each point is described by a SIFT descriptor [9] for matching. Fig. 2 shows examples of the computed registrations.

Guided Feature Matching. Let I_q and I_t be two images to register and T_k the current transform estimate between I_q and I_t . The subscripts i and j indicate matching features in images I_q and I_t with locations \mathbf{x}_i , \mathbf{y}_j and descriptor vectors Ψ_i and Ψ_j respectively. Feature point matching is formulated as a linear assignment problem with unary costs C_{ij} defined as:

$$C_{ij} = \begin{cases} +\infty & \text{if } C_{ij}^{geom} > r \\ w^{desc} C_{ij}^{desc} + w^{geom} C_{ij}^{geom} & \text{otherwise.} \end{cases} \quad (1)$$

It depends on the descriptors distance $C_{ij}^{desc} = \|\Psi_i - \Psi_j\|_2$ as well as the symmetric transfer error $C_{ij}^{geom} = \|T_k(\mathbf{x}_i) - \mathbf{y}_j\|_2 + \|\mathbf{x}_i - T_k^{-1}(\mathbf{y}_j)\|_2$. The hard threshold r on C_{ij}^{geom} allows matching only within a spatial neighbourhood of a feature. This increases matching robustness while reducing computational complexity.

Robust Thin Plate Spline Estimation. Direct TPS computation based on all feature point matches computed at the previous step leads to inaccuracies due to occasional mismatches. To filter them out we employ the LO-RANSAC [2] framework. In our implementation two transformation models of different complexity are utilised for hypothesis testing. A similarity transform with a *loose* threshold is used for fast initial outlier rejection, while a TPS is fitted only to the inliers of the few promising hypotheses. The resulting TPS warp T_{k+1} is the one with the most inliers.

ROI Localisation Refinement. Given an ROI in the query image we wish to obtain the corresponding ROI in the target, i.e. the ROI covering the same “object”. The TPS transform T registering the query and target images provides a rough estimate of the target ROI as a quadrilateral R_t^0 which is a warp of the query rectangle R_q . However, possible inaccuracies in T may cause R_t^0 to be misaligned with the actual ROI, and in turn this may hamper ROI ranking. To alleviate this problem, the detected ROI can be fine-tuned by locally maximizing the normalised intensity cross-correlation between the query rectangle and the target quadrilateral. This task is formulated as a constrained non-linear least squares problem where each vertex is restricted to a box to avoid degeneracies. An example is shown in Fig. 4.

2.3 Scalable Registration by Transform Composition

When adding a new image to the repository, it has to be registered with all target images (images of the same class). A naïve implementation results in the

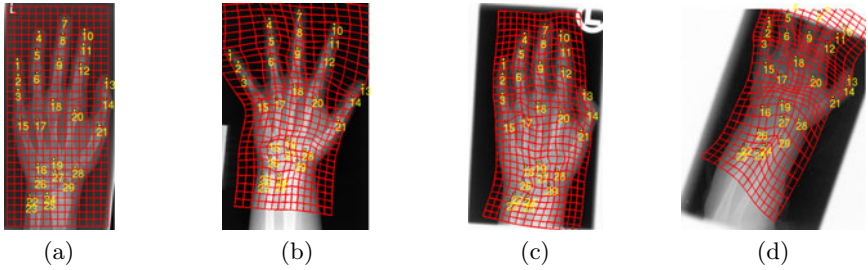


Fig. 2. Robust thin plate spline matching. (a): query image with a rectangular grid and a set of ground-truth (GT) landmarks (shown with yellow numbers); (b)-(d): target images showing the GT points mapped via the *automatically* computed transform (GT points not used) and the induced grid deformation

computational complexity linear in the number of target images, which grows together with the repository size and quickly becomes infeasible as image registration is computationally heavy. Instead, we use a scalable technique suitable for large datasets. A new image q is registered with only a *fixed* subset of E target images (exemplars), which results in E transforms $T_{q,e}$, $e = 1 \dots E$. The transformations $T_{e,t}$ between an exemplar e and each of the remaining target set images t are pre-computed. Then the transformation between images q and t can be obtained by transform composition (using different exemplars) followed by robust aggregation as $T_{q,t} = \text{median}_e (T_{q,e} \circ T_{e,t})$. While the complexity is still linear in the number of target images, its advantage is that only $E = \text{const}$ registrations should be computed, while transform composition is a cheap operation. The technique is related to the multi-atlas segmentation scheme of [6]. From our experiments (not presented due to the space restrictions), the accuracy of exemplar-based registration is similar to the pairwise case.

2.4 ROI Ranking Functions

At this stage we have obtained ROIs in a set of target images, corresponding to the ROI in the query image. The question then remains of how to order the images for the retrieval system, and this is application dependent. We consider three choices of the ranking function defined as the similarity $S(I_q, R_q, I_t, R_t)$ between the query and target ROIs, R_q, R_t and images I_q, I_t . The retrieval results are ranked in decreasing order of S . The similarity S can be defined to depend on the ROI Appearance (ROIA) only. For instance, the normalised cross-correlation (NCC) of ROI intensities can be used. The S function can be readily extended to accommodate the ROI Shape (ROISA) as $S = (1 - w) \min(E_q, E_t) / \max(E_q, E_t) + w \text{NCC}(R_q, R_t)$, where E_q and E_t are elongation coefficients (ratio of major to minor axis) of query and target ROIs, and $w \in [0, 1]$ is a user tunable parameter. At the other extreme, the function S can be tuned to capture global Image Geometry (IG) cues. If similar scale scans are of interest, then S can be defined as: $S(I_q, R_q, I_t, R_t) = (1 - w) \min\{\Sigma, 1/\Sigma\} +$







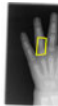



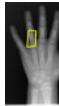

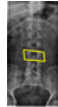


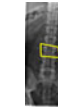
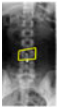
Query image and ROI	Ranking function	Top-5 retrieved images with detected ROI				
	IG ($w = 0.5$)					
	ROISA ($w = 0.5$)					
	ROIA					

Fig. 3. The effect of different ranking functions on ROI retrieval. ROIs are shown in yellow. IG retrieves scans with similar image cropping; ROISA ranks paediatric hands high because the query is paediatric; ROIA ranks based on ROI intensity similarity.

$w \text{NCC}(R_q, R_t)$, where $\Sigma > 0$ is the scale of the similarity transform computed from feature point matches, and $w \in [0, 1]$ is a user tunable parameter.

Fig. 3 shows the top ranked images retrieved by these functions. This is an example of how local ROI clues can be employed for ranking, which is not possible with global, image-level visual search. In clinical practice, ranking functions specifically tuned for a particular application could be used, e.g. trained to rank on the presence of a specific anomaly (e.g. nodules or cysts).

3 Results and Comparisons

The dataset. The dataset contains X-ray images of five classes: **hand**, **spine**, **chest**, **cranium**, **negative (the rest)** taken from the publicly available IRMA dataset [3]. Each class is represented by 205 images. The negative class contains images of miscellaneous body parts not included in the other classes. The images are stored in the PNG format without any additional textual metadata. Images within each class exhibit a high amount of variance, e.g. scale changes, missing parts, new objects added (overlaid writings), anatomy configuration changes (phalanges apart or close to each other). Each of the classes is randomly split into 65 testing, 70 training, and 70 validation images.

Image classification performance is measured by the ratio of correctly classified test images to the total number of test images. The overall accuracy is 98%. The few misclassifications are caused by the overlap between the **negative** class and other classes, if the negative image partially contains the same body part.

Accuracy of Structured Image Retrieval. To evaluate the accuracy of ROI retrieval from the dataset, we annotated test hand and spine images with

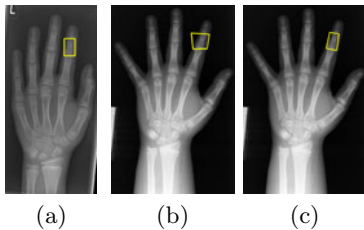


Fig. 4. (a): query; (b),(c): target before and after local refinement

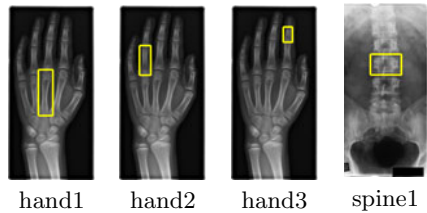


Fig. 5. Four annotated bones used for the retrieval performance assessment

Table 1. Comparison of image retrieval accuracy

Method	hand1		hand2		hand3		spine1	
	meanAP	medAP	meanAP	medAP	meanAP	medAP	meanAP	medAP
Proposed	0.81	0.89	0.85	0.90	0.65	0.71	0.49	0.51
Baseline	0.68	0.71	0.66	0.71	0.38	0.36	0.35	0.35
<code>elastix</code>	0.62	0.67	0.61	0.68	0.38	0.37	0.22	0.19

axis-aligned bounding boxes around the same bones as shown in Fig. 5. The ROI retrieval evaluation procedure is based on that of PASCAL VOC detection challenge [4]. A query image and ROI are selected from the test set and the corresponding ROIs are retrieved from the rest of the test set using the proposed algorithm. A detected ROI quadrangle is labelled as correct if the overlap score between its axis-aligned bounding box and the ground truth one is above a threshold. The retrieval performance for a query is assessed using the Average Precision (AP) measure computed as the area under the “precision vs recall” curve. Once the retrieval performance is estimated for each of the images as a query, its mean (meanAP) and median (medAP) over all queries are taken as measures. We compare the retrieval performance of the framework (ROIA ranking, no ROI refinement) using different registration methods: the proposed one (Sect. 2.2), baseline feature matching with affine transform [11], and `elastix` B-splines [7]. All three methods compute pairwise registration (i.e. no exemplars).

The proposed algorithm outperforms the others on all types of queries (Table 1). As opposed to the baseline, our framework can capture non-rigid transforms; intensity-based non-rigid `elastix` registration is not robust enough to cope with the diverse test set. Compared to hand images, worse performance on the spine is caused by less consistent feature detections on cluttered images.

Complexity and Speed. The retrieval framework is efficient and scalable, which allows for the immediate ROI retrieval (in a fraction of second) even using our current non-optimised Matlab implementation. Image class information as well as pairwise registrations between same-class images are pre-computed and stored for the repository images, so they are immediately available if a query image is from the repository. The complexity of adding a new image splits into the following parts: (i) image classification has constant complexity ($\approx 0.5s$ per

image); (ii) registration with a *fixed* number of target set exemplars has constant complexity ($\approx 2s$ per exemplar); (iii) registration with the rest of the target images is linear in the number of images, but the transform composition is very quick compared to registration. Therefore, the most computationally complex operations are invoked only a fixed number of times.

4 Conclusion

We have presented a new framework for the immediate retrieval of medical images and simultaneous, automatic localisation of anatomical structures of interest. Robustness with respect to repeated structures is incorporated via non-rigid image registration driven by guided robust sparse feature matching. Supervised image-level classification also contributes to the high level of accuracy demonstrated on a publicly available labelled database of X-ray images. The proposed visual search framework is fairly generic and can be extended to different modalities/dimensionalities with a proper choice of intra-class registration methods. An interactive demo of the ROI retrieval framework is available on http://www.robots.ox.ac.uk/~vgg/research/med_search/

References

1. Chui, H., Rangarajan, A.: A new point matching algorithm for non-rigid registration. *Comp. Vis. and Image Understanding* 89(2-3), 114–141 (2003)
2. Chum, O., Matas, J., Obdržálek, Š.: Enhancing RANSAC by generalized model optimization. In: *Proc. of Asian Conf. on Comp. Vis.*, vol. 2, pp. 812–817 (2004)
3. Deserno, T.M.: IRMA dataset (2009), http://ganymed.imib.rwth-aachen.de/irma/datasets_en.php?SELECTED=00009#00009.dataset
4. Everingham, M., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A.: The PASCAL Visual Object Classes (VOC) challenge. *Int. J. Comp. Vis.* 88(2), 303–338 (2010)
5. Guld, M.O., Kohnen, M., Keyzers, D., Schubert, H., Wein, B., Bredno, J., Lehmann, T.M.: Quality of DICOM header information for image categorization. In: *Int. Symp. Med. Imag. SPIE*, vol. 4685, pp. 280–287 (February 2002)
6. Isgum, I., Staring, M., Rutten, A., Prokop, M., Viergever, M., van Ginneken, B.: Multi-atlas-based segmentation with local decision fusion – application to cardiac and aortic segmentation in CT scans. *IEEE Trans. Med. Imag.* 28(7), 1000–1010 (2009)
7. Klein, S., Staring, M., Murphy, K., Viergever, M.A., Pluim, J.P.W.: elastix: A toolbox for intensity-based medical image registration. *IEEE Trans. Med. Imag.* 29(1), 196–205 (2010)
8. Lam, M., Disney, T., Pham, M., Raicu, D., Furst, J., Susomboon, R.: Content-based image retrieval for pulmonary computed tomography nodule images. In: *Med. Imag. 2007: PACS and Imag. Inform. SPIE*, vol. 6516 (March 2007)
9. Lowe, D.: Distinctive image features from scale-invariant keypoints. *Int. J. Comp. Vis.* 60, 91–110 (2004)

10. Müller, H., Michoux, N., Bandon, D., Geissbuhler, A.: A review of content-based image retrieval systems in medical applications – clinical benefits and future directions. *Int. J. Med. Inform.* 73(1), 1–23 (2004)
11. Philbin, J., Chum, O., Isard, M., Sivic, J., Zisserman, A.: Object retrieval with large vocabularies and fast spatial matching. In: *Proc. IEEE Conf. on Comp. Vis. and Patt. Recog.* (2007)
12. Sivic, J., Zisserman, A.: Efficient visual search of videos cast as text retrieval. *IEEE Trans. Patt. Anal. and Mach. Intell.* 31(4), 591–606 (2009)
13. Vedaldi, A., Gulshan, V., Varma, M., Zisserman, A.: Multiple kernels for object detection. In: *Proc. Int. Conf. Comp. Vis.*, pp. 606–613 (2009)

Retrieval Evaluation and Distance Learning from Perceived Similarity between Endomicroscopy Videos

Barbara André^{1,2}, Tom Vercauteren¹, Anna M. Buchner³,
Michael B. Wallace⁴, and Nicholas Ayache²

¹ Mauna Kea Technologies, Paris

² INRIA - Asclepios, Sophia-Antipolis

³ Hospital of the University of Pennsylvania, Philadelphia

⁴ Mayo Clinic, Jacksonville, Florida

Abstract. Evaluating content-based retrieval (CBR) is challenging because it requires an adequate ground-truth. When the available ground-truth is limited to textual metadata such as pathological classes, retrieval results can only be evaluated indirectly, for example in terms of classification performance. In this study we first present a tool to generate perceived similarity ground-truth that enables direct evaluation of endomicroscopic video retrieval. This tool uses a four-points Likert scale and collects subjective pairwise similarities perceived by multiple expert observers. We then evaluate against the generated ground-truth a previously developed dense bag-of-visual-words method for endomicroscopic video retrieval. Confirming the results of previous indirect evaluation based on classification, our direct evaluation shows that this method significantly outperforms several other state-of-the-art CBR methods. In a second step, we propose to improve the CBR method by learning an adjusted similarity metric from the perceived similarity ground-truth. By minimizing a margin-based cost function that differentiates similar and dissimilar video pairs, we learn a weight vector applied to the visual word signatures of videos. Using cross-validation, we demonstrate that the learned similarity distance is significantly better correlated with the perceived similarity than the original visual-word-based distance.

1 Introduction

Successfully developed in the field of computer vision, content-based retrieval (CBR) methods also have valuable applications in the field of medical imaging. In particular, probe-based confocal laser endomicroscopy (pCLE) is a recent imaging technology that enables the endoscopist to acquire, *in vivo*, microscopic video sequences of the epithelium. Because *in vivo* diagnostic interpretation of a pCLE video is still challenging for many endoscopists, it could be supported by the automated real-time extraction of visually similar videos that have already been annotated with a textual diagnosis. We previously developed in [1] a dense bag-of-visual-words (BoW) method for pCLE video retrieval, called “Dense-Sift”, which

provides qualitatively relevant retrieval results. When evaluated in terms of classification, “Dense-Sift” was shown to significantly outperform several state-of-the-art CBR methods referred to as “competitive methods” in Section 2. However, there is a high variability in the appearance of pCLE videos, even within the same pathological class. In order to measure the adequacy of pCLE video retrieval, we propose to evaluate the “Dense-Sift” method directly in terms of visual similarity distance. To this purpose, we develop in Section 3 an online survey tool called “Visual Similarity Scoring” (VSS) to help pCLE experts in generating a perceived similarity ground-truth. In Section 4, all state-of-the-art methods are evaluated against the generated ground-truth and we show that, with statistical significance, “Dense-Sift” proves to be the best. Leveraging the perceived similarity used for evaluation purposes, we propose, in a second step, to improve the “Dense-Sift” retrieval method by learning from this ground truth an adjusted similarity metric. Our metric learning technique, presented in Section 5, is based on a visual word weighting scheme which we evaluate using cross-validation. The learned similarity metric is shown to be significantly better correlated with the perceived similarity than the original visual-word-based distance.

2 State-of-the-Art in CBR and Distance Metric Learning

Among the state-of-the-art methods in CBR, the BoW method of Zhang et al. [2], referred to as “HH-Sift”, is particularly successful for the retrieval of texture images in computer vision. Whereas “HH-Sift” combines the sparse “Harris-Hessian” detector with the SIFT descriptor, the competitive “Textons” method proposed by Leung and Malik [3] is based on a dense description of local texture features. Adjusting these approaches for pCLE retrieval, the “Dense-Sift” method of [1] uses a dense SIFT description. Such a description is invariant to in-plane rotations or translations changes that are due to the motion of the pCLE miniprobe in contact with the epithelium, and to the possible illumination changes that are due to the leakage of fluorescein used in pCLE. “Dense-Sift” also enables the extension from pCLE image retrieval to pCLE video retrieval by leveraging video mosaicing results. Another CBR method shown as a competitive method in [1] is the “Haralick” [4] method based on global statistical features. In this study, we choose to evaluate these four CBR methods in terms of visual similarity distance, in order to compare their retrieval performances.

Distance metric learning has been investigated by rather recent studies to improve classification or recognition methods. Yang et al. [5] proposed a “boosted distance metric learning” method that projects images into a Hamming space where each dimension corresponds to the output of a weak classifier. Weinberger et al. [6] explored convex optimizations to learn a Mahalanobis transformation such that distances between nearby images are shrunk if the images belong to the same class and expanded otherwise. At the level of image descriptors, Philbin et al. [7] have a similar approach that transforms the description vectors into a space where the clustering step more likely assigns matching descriptors to the same visual word and non-matching descriptors to different visual words.

Since the second approach relies on a matching ground-truth that is closer to the pairwise similarity ground-truth that we present in the next section, our proposed metric learning technique is inspired from the method of [7] and applies the transformation to the visual words signatures of videos.

3 Generation of Perceived Similarity Ground-Truth

Our video database contains 118 pCLE videos of colonic polyps that were acquired from 66 patients for the study of Buchner et al. [8]. The length of these videos is ranging from 1 seconds to 4 minutes. To generate a pairwise similarity ground-truth between these videos, we designed an online survey tool, called VSS [9], that allows multiple human observers, who are fully blinded to the video metadata such as the pCLE diagnosis, to qualitatively estimate the perceived visual similarity degree between videos. The VSS tool proposes, for each video couple, the following four-points Likert scale: “very dissimilar”, “rather dissimilar”, “rather similar” and “very similar”. Because interpreting whole video sequences requires a lot of time, the VSS supports this task by making available the whole video content and for each video, a set of static mosaic images providing a visual summary. Indeed, Dabizzi et al. [10] recently showed that pCLE mosaics have the potential to replace pCLE videos for a comparable diagnosis accuracy and a significantly shorter interpretation time. We also paid attention to how video couples should be drawn by the VSS. If the video couples had been randomly drawn, the probability of drawing dissimilar videos would be much higher than the probability of drawing very similar videos, which would thus be poorly represented in ground-truth data. To solve this problem, we used the *a priori* similarity distance D_{prior} computed by the “Dense-Sift” method to

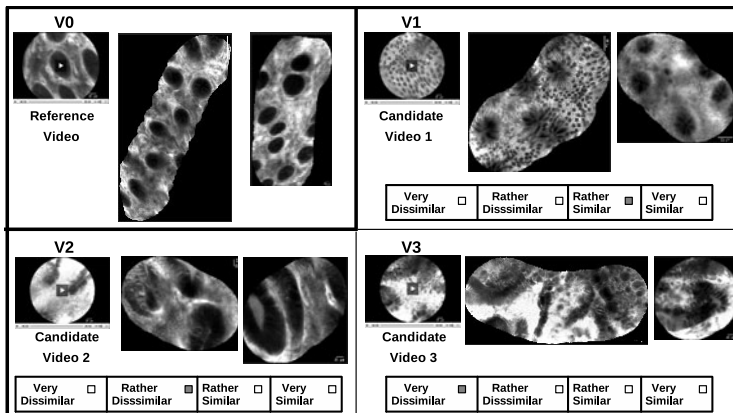


Fig. 1. Schematic outline of the online “Visual Similarity Scoring” tool showing an example of a scoring process, where 3 video couples ($V0, V1$), ($V0, V2$) and ($V0, V3$) are proposed. Each video is summarized by a set of mosaic images.

draw according to the following law: the probability of drawing a video couple $(V1, V2)$ is proportional to the inverse of the density of $D_{prior}(V1, V2)$. Each scoring process, as illustrated in Fig. 1, is defined by the drawing of 3 video couples $(V0, V1)$, $(V0, V2)$ and $(V0, V3)$, where the candidate videos $V1, V2$ and $V3$ belong to patients that are different from the patient of the reference video $V0$, in order to exclude any patient-related bias. 17 observers, ranging from middle expert to expert in pCLE diagnosis, performed as many scoring processes as they could. The averaging time to score 3 video couples during one scoring process was 10 minutes. Our generated ground-truth can be represented as an oriented graph $G = (V, E)$ where the nodes V are the videos and where each couple of videos may be connected by zero, one or multiple edges representing the similarity scores. As less than 1% of these video couples were scored by more than 4 distinct observers, it was not relevant to measure inter-observer variability. In total, 3804 similarity scores were given on 1951 distinct video couples. Only 14.5% of all 13434 distinct video couples were scored. Although the similarity graph is very sparse, we demonstrate in the following sections that it constitutes a first valuable ground-truth, not only for retrieval evaluation but also to learn an adjusted similarity distance.

4 Evaluation of CBR Methods against Ground-Truth

The evaluation of a CBR method against ground-truth can be qualitatively illustrated by the four superimposed histograms $H_L, L \in \{-2, -1, +1, +2\}$ shown in

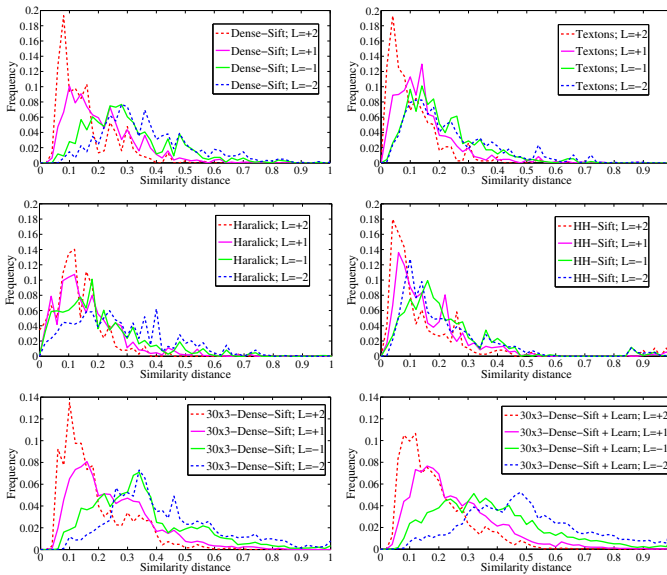


Fig. 2. Histograms of L -scored similarity distances computed by the CBR methods

Fig. 2: H_L is the histogram of the similarity distances which were computed by the CBR method in the restricted domain of all L -scored video couples, where L is one of the four Likert points from “very dissimilar” to “very similar”. We observe that these histograms are correctly ordered with respect to the four Likert points for all methods, except for “HH-Sift” that switches H_{-1} and H_{-2} . We also notice that the histograms are better separated for “Dense-Sift” than for the other methods. This is quantitatively confirmed by the histogram separability measures, given by the Bhattacharyya distance, that are shown in the supplemental material at <http://hal.inria.fr/inria-00598301/fr/>.

Possible indicators of the correlation between the visual-word-based similarity distance and the ground-truth distance are Pearson π product moment, Spearman ρ and Kendall τ . Compared to π which measures linear dependence based on the data values, ρ and τ are better adapted to the psychometric Likert scale because they measure monotone dependence based on the data ranks [11]. Kendall τ is less commonly used than Spearman ρ but its interpretation in terms of probabilities is more intuitive. To assess statistical significance for the comparison between the correlation coefficients that are associated to each CBR method, we have to perform the adequate statistical test. First, ground-truth data lying on the four-points Likert scale are characterized by a non-normal distribution, so data ranks should be used instead of data values. Second, the rank correlation coefficients measured for two methods are themselves correlated because they both depend on the same ground-truth data. For these reasons, we perform Steiger’s Z -tests, as recommended by Meng et al. [12], and we apply it to Kendall τ . The correlation results shown in Table 1 demonstrate that, with statistical significance, “Dense-Sift” is better than all other competitive methods, while “Textons” and “Haralick” are better than “HH-Sift”.

Standard recall curves are a common means of evaluating retrieval performance. However, because of the sparsity of the ground-truth, it is not possible to compute them in our case. Instead, we need “sparse recall” curves. At a fixed number k , we define the sparse recall value of a CBR method as the percentage of L -scored video couples, with $L = +1$ or $+2$, for which one of the two videos has

Table 1. Indicators of correlation between similarity distance computed by the CBR methods and ground-truth. σ_{est} is the standard deviation of the estimator, it can be computed from the standard deviation of the n samples $\sigma_{samples} = \sqrt{n-1} \cdot \sigma_{est}$. The difference between methods **M4** and **M5** is not statistically significant ($p > 0.3$).

CBIR method M	M1 30x3-DS-learn	M2 30x3-DS	M3 Dense-Sift (DS)	M4 Textons	M5 Haralick	M6 HH-Sift
Pearson π	52.8 %	46.0 %	47.8 %	32.7 %	33.8 %	15.7 %
standard error σ_{est}	0.8 %	0.9 %				
Spearman ρ	56.6 %	49.0 %	51.5 %	35.4 %	34.2 %	21.8 %
standard error σ_{est}	0.9 %	1.1 %				
Kendall τ	52.6 %	45.2 %	47.0 %	32.1 %	30.6 %	19.4 %
standard error σ_{est}	0.9 %	1.0 %				
Steiger’s Z -test on τ	> M2		> M4-M5-M6	> M6	> M6	
p-value p	$p = 0.018$		$p < 10^{-4}$	$p < 10^{-4}$	$p < 10^{-4}$	

been retrieved among the k nearest neighbors of the other video. Sparse recall curves in Fig. 3 show that “Dense-Sift” extracts similar videos much faster than the other methods in a small retrieval neighborhood, which is clinically relevant for our pCLE application. Thus, local similarity distances are better captured by the “Dense-Sift” method.

5 Distance Learning from Perceived Similarity

As mentioned in Section 2, we now propose a metric learning technique that is inspired from the method of Philbin et al. [7]. Our objective is to transform video signatures, that are histograms of visual words, into new signatures where visual words are now weighted by a vector w that better discriminates between similar videos and dissimilar videos. We thus consider two groups: D_+ is the set of video couples that have been scored with +2 or +1, and D_- is the set of video couples that have been scored with -2 or -1. Our constraints are the following: the weights w should be positive in order to maintain the positiveness of visual word frequencies, the χ^2 metric used by standard BoW methods should

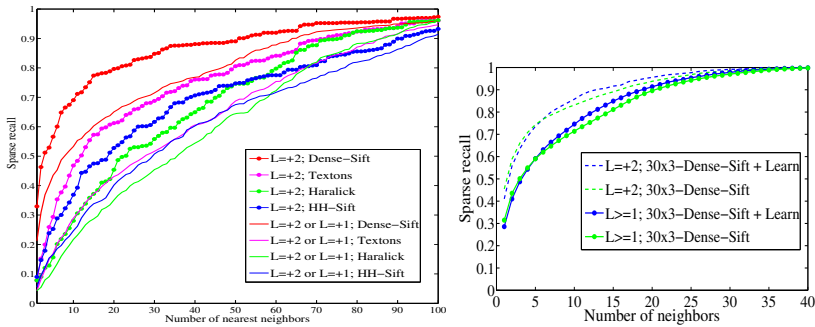


Fig. 3. Sparse recall curves associated the two methods in L -scored domains

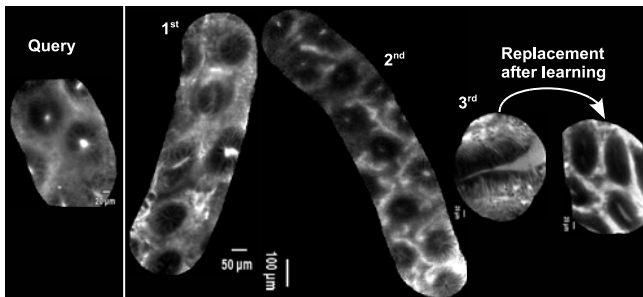


Fig. 4. Example of pCLE video query with its 3 nearest neighbors retrieved by “Dense-Sift” before and after metric learning. Videos are represented by mosaic images.

be the distance between video signatures and the new signatures should be $L1$ -normalized before χ^2 distances are measured. We optimize the transformation w by minimizing the following margin-based cost function:

$$f(w) = \frac{1}{\text{Card}(D_+)} \sum_{(x,y) \in D_+} L(b - \chi^2(\frac{w \circ s_x}{\|w \circ s_x\|_{L1}}, \frac{w \circ s_y}{\|w \circ s_y\|_{L1}})) + \gamma \frac{1}{\text{Card}(D_-)} \sum_{(x,y) \in D_-} L(\chi^2(\frac{w \circ s_x}{\|w \circ s_x\|_{L1}}, \frac{w \circ s_y}{\|w \circ s_y\|_{L1}}) - b) \quad (1)$$

where b is the margin, s_x is the visual word signature of the video x , \circ is the Hadamard (element-wise) product, $L(z) = \log(1 + e^{-z})$ is the logistic-loss function and γ is a constant that potentially penalizes either dissimilar nearby videos or similar remote videos. The learned similarity distance is then defined by:

$$D_{learn}(x, y) = \chi^2(\frac{w_{opt} \circ s_x}{\|w_{opt} \circ s_x\|_{L1}}, \frac{w_{opt} \circ s_y}{\|w_{opt} \circ s_y\|_{L1}}) \quad (2)$$

To exclude the learning bias, we apply this distance learning technique using $m \times q$ -fold cross-validation: we performed m random partitions of our database into q video subsets. Each of these subsets is successively the testing set and the two others the training set for both video retrieval and distance learning. To eliminate patient-related bias, all videos from the same patient are in the same subset. Given our sparse ground-truth, q must be not too large in order to have enough similarity scores in each testing set and not too small to ensure enough similarity scores in the training set.

For our experiments, we took $m = 30$ and $q = 3$. Then, by choosing $\gamma = 5$ and $b = (\text{median}(H_{+2,+1}^{train}) + \text{median}(H_{-2,-1}^{train}))/2$ as an intuitive value for the margin b for each training set, we show in the following that we obtain satisfying correlation results with respect to the ground truth.

As ‘‘Dense-Sift’’ proved to be the best CBR method, we propose to use its visual word signatures as inputs of the learning process in order to improve its visual-word-based distance. In order to compare the performances of the learned similarity distance with those of the visual-word-based distance, ‘‘Dense-Sift’’ was re-trained on each training subset and re-evaluated on corresponding testing subsets. We call ‘‘30x3-fold-Dense-Sift’’ the cross-validated ‘‘Dense-Sift’’ without metric learning and ‘‘30x3-fold-Dense-Sift-learn’’ the same one improved by metric learning. The superimposed histograms H_L for the retrieval method before and after learning are represented in the bottom of Fig. 2. We observe that these histograms are better separated after the metric learning process, which is confirmed by the Bhattacharyya distances shown in the supplemental material.

Although the sparse recall curves of the retrieval method before and after learning are very close to each other, as shown in Fig. 3, the metric learning process globally improved the performance of the retrieval method in terms of perceived visual similarities. Indeed, the correlation results shown in Table 1 demonstrate that, with statistical significance, the learned similarity distance is better correlated than the original visual-word-based distance with the ground-truth similarity distance. Besides, for some cases as the one shown in Fig. 4, we observe that first neighbors are qualitatively more similar after metric learning.

6 Conclusion

The main contributions of this study are the generation of a valuable ground-truth for perceived visual similarities between endomicroscopic videos, the evaluation of content-based retrieval methods in terms of visual similarity and the learning of an adjusted similarity distance. The proposed methodology could be applied to other medical or non-medical databases. Our evaluation experiments confirmed that the dense BoW method for endomicroscopic video retrieval has better performances than other competitive methods, not only in terms of pathological classification but also in terms of correlation with a ground-truth similarity distance. Future work will focus on enlarging the ground truth database to investigate more sophisticated metric learning techniques. Our long-term goal is to improve endomicroscopic video retrieval and assess whether it could support the endoscopists in establishing more accurate *in vivo* diagnosis.

References

1. André, B., Vercauteren, T., Wallace, M.B., Buchner, A.M., Ayache, N.: Endomicroscopic video retrieval using mosaicing and visual words. In: Proc. ISBI 2010, pp. 1419–1422 (2010)
2. Zhang, J., Lazebnik, S., Schmid, C.: Local features and kernels for classification of texture and object categories: a comprehensive study. *Int. J. Comput. Vis.* 73, 213–238 (2007)
3. Leung, T., Malik, J.: Representing and recognizing the visual appearance of materials using three-dimensional textons. *Int. J. Comput. Vis.* 43, 29–44 (2001)
4. Haralick, R.M.: Statistical and structural approaches to texture. *Proc. IEEE* 67, 786–804 (1979)
5. Yang, L., Jin, R., Mummert, L., Sukthankar, R., Goode, A., Zheng, B., Hoi, S.C.H., Satyanarayanan, M.: A boosting framework for visuality-preserving distance metric learning and its application to medical image retrieval. *IEEE Trans. Pattern Anal. Mach. Intell.* 32, 30–44 (2010)
6. Weinberger, K.Q., Saul, L.K.: Distance metric learning for large margin nearest neighbor classification. *J. Mach. Learn. Res.* 10, 207–244 (2009)
7. Philbin, J., Isard, M., Sivic, J., Zisserman, A.: Descriptor learning for efficient retrieval. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) *ECCV 2010, Part III*. LNCS, vol. 6313, pp. 677–691. Springer, Heidelberg (2010)
8. Buchner, A.M., Shahid, M.W., Heckman, M.G., Krishna, M., Ghabril, M., Hasan, M., Crook, J.E., Gomez, V., Raimondo, M., Woodward, T., Wolfsen, H., Wallace, M.B.: Comparison of probe based confocal laser endomicroscopy with virtual chromoendoscopy for classification of colon polyps. *Gastroenterology* 138(3), 834–842 (2009)
9. Visual Similarity Scoring (VSS), <http://smartatlas.maunakeatech.com>, login: MICCAI-User, password: MICCAI2011
10. Dabizzi, E., Shahid, M.W., Qumseya, B., Othman, M., Wallace, M.B.: Comparison between video and mosaics viewing mode of confocal laser endomicroscopy (pCLE) in patients with barrett's esophagus. *Gastroenterology (DDW 2011)* (2011)
11. Barnett, V.: *Sample survey principles and methods* (1991)
12. Meng, X.L., Rosenthal, R., Rubin, D.B.: Comparing correlated correlation coefficients. *Psychological Bulletin* 111(1), 172–175 (1992)

A New Approach for Tubular Structure Modeling and Segmentation Using Graph-Based Techniques

Jack H. Noble and Benoit M. Dawant

Dept. of Electrical Engineering and Computer Science, Vanderbilt University
Nashville, TN 37235, USA
{jack.h.noble, benoit.dawant}@vanderbilt.edu

Abstract. In this work, a new approach for tubular structure segmentation is presented. This approach consists of two parts: (1) automatic model construction from manually segmented exemplars and (2) segmentation of structures in unknown images using these models. The segmentation problem is solved by finding an optimal path in a high-dimensional graph. The graph is designed with novel structures that permit the incorporation of prior information from the model into the optimization process and account for several weaknesses of traditional graph-based approaches. The generality of the approach is demonstrated by testing it on four challenging segmentation tasks: the optic pathways, the facial nerve, the chorda tympani, and the carotid artery. In all four cases, excellent agreement between automatic and manual segmentations is achieved.

Keywords: Tubular segmentation, optimal paths, Dijkstra's algorithm.

1 Introduction

Tubular structure localization has been studied extensively over the past few decades (see Lesage et al. [1] for an extensive review). One common segmentation approach is to use anatomy-specific models constructed from a training set. For instance, de Bruijne et al. [2] use specialized active shape models to localize abdominal aortic aneurysms. In the work presented herein, a new algorithm is presented that can be used to segment a wide variety of tubular structures. The method uses a novel approach to build anatomy-specific models and only requires a small number of sample segmentations for training. The segmentation problem is solved by finding an optimal path in a high-dimensional graph. The graph is designed with novel structures that permit the incorporation of prior information from the model into the optimization and account for several weaknesses of traditional graph-based approaches.

One common approach for tubular structure extraction is to treat it as the problem of finding an optimal path in a graph, $G=\{V,E\}$ [3]. Typically, when used for tubular structure segmentation, V , the set of nodes, corresponds to the image voxels; E the set of edges, connects neighboring voxels; and the costs associated with the edges are defined by some heuristically derived function. This basic approach is widely used because it is computationally efficient and finds a global optimum. The algorithm also

has four identifiable limitations: (1) It is difficult to incorporate local and/or training-based prior knowledge into the segmentation framework. Thus, the SOI (Structure Of Interest) must be enhanced prior to segmentation. This is not always possible as many tubular structures have intensity characteristics that vary along their length and borders that lack contrast. (2) The algorithm identifies a curve, rather than the 3D structure. Further processing must be applied to identify the full structure (an alternative is to add a 4th dimension corresponding to width as done in [4]). (3) The algorithm is biased with respect to length, i.e., because the algorithm finds the path composed of edges with the minimum cumulative costs, it tends to favor shorter, straight paths, which can result either in segmentation failure or in erroneous shortcutting at structure corners. (4) The method requires manual definition and tuning of cost functions.

The algorithm presented in this paper uses an unconventional optimal path-based extraction scheme. Segmentation is performed on a unique high-dimensional graph that is designed to incorporate information from a novel tubular structure model. This approach addresses limitations of the traditional 3D formulation.

2 Methods

Although other features could be used, in this work intensity and shape-based features that are used for training and segmentation include: (1) the curve orientation, measured using finite differences, which provides a local prior on the curve's shape; (2) intensity values along radii normal to the curve at distances of 0, 0.2, 0.9, and 1.1 times the width of the structure at every $\pi/4$ radians; and (3) intensity derivatives on those radii at 0.1 and 1.0 times the structure width. The radii directions chosen when building the model are stored and used to compute the features on the same radii in target images for segmentation.

2.1 The Tubular Structure Model and Its Training

In our approach, a large number of features are used, but techniques will be used to weigh the features that are locally the most discriminant. The approach relies on separating the SOI into a set of segments, which will be referred to as "localities" $\{l\}$. Each locality will correspond to the same anatomic substructures across samples. The separation of the SOI into localities is done automatically as part of the training algorithm. A locality l of a particular training sample s can range from one to any number of voxels in length. Figure 1 shows a synthetic 2D example of a SOI with 2 training samples divided into 5 localities. As shown, the localities correspond to similar regions across each training structure. The model training procedure is completed in four steps: (1) point correspondence across training centerlines is established, (2) the boundaries of the localities are determined, (3) model features are computed, and (4) weights for each feature are established.

Point correspondence across centerlines is necessary to determine the localities that correspond to the same distinct regions across the sample SOIs, and it is extracted using a graph-based optimization. One training volume is chosen as the reference.

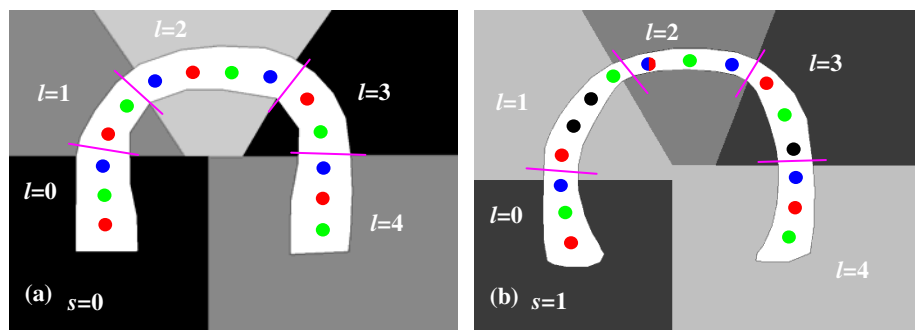


Fig. 1. Synthetic 2D example of a structure model. (a) and (b) are the reference and training SOIs. RGB values of centerline points indicate correspondence across volumes. Locality divisions are indicated by the purple lines.

The search determines one voxel along each training centerline to be associated with each reference centerline voxel. The cost function used in the search is optimized when local features between the corresponding voxels are best matched. The color-coded voxels in the centerlines in Figure 1 represent example correspondences. In (b), black voxels indicate no association with a reference voxel, and multiple colors indicate multiple associations. The main purpose of this step is to provide a way to establish a correspondence between locality boundaries across training samples. Thus, a one-to-one correspondence between centerline points is not necessary.

Although this could be automated, the number of localities for each SOI is chosen manually. However, the positions of the locality boundaries are determined automatically using a gradient descent technique. The cost function used favors smaller intra-locality and larger inter-locality variance of model features across all the training samples, and it includes a regularization term that favors boundaries spread evenly across the SOI. An example of optimal locality selection is shown in Figure 1. Once the localities are determined, the mean and variance of each feature for each locality/sample are computed.

The next step involves determining the features that are the most discriminant in each locality. A desirable feature is one that has a low value inside the SOI relative to the background. To identify these, features are first computed across the training images and the resulting feature images are convolved with a Gaussian-based valley detecting filter with scale equal to the structure width. A high positive response indicates a desirable feature. A low or negative response indicates an undesirable one. Weights for each feature are thus set to the response of the filter if it is positive and set to zero otherwise. After this step the creation of the models is completed. These consist of a sequence of structure samples, each sectioned into localities representing similar anatomical regions; and statistical values and importance weightings for a set of features within each sample/locality.

2.2 Structure Segmentation

Once the structure model is constructed, it can be used to segment new images using an optimal path finding algorithm. This algorithm requires three pieces of information: (1) the curve's starting and ending voxels, (2) a defined graph structure, and (3) a defined cost function. In this work, the starting and ending voxels were chosen manually (this could be automated in future work using, for instance, atlas-based methods as was done in [5,7]).

To achieve our goal, a novel graph structure has been implemented that is unique in two ways: (1) Nodes are associated with specific training samples/localities. This allows costs to be computed using the model's features associated with each sample/locality. Also, this permits reconstruction of the full 3D structure from the optimal path by assuming, since the features are width dependent, that the width of the structure at a node in the optimal path should match the width of the sample in the locality indicated by the node. (2) Novel graph structures are implemented such that the only paths that exist between the seed and end-node obey hard constraints on the minimum/maximum length of the structure. The length of a path in a graph can be constrained by creating an abstract dimension P in the graph. The seed lies at $p=0$ in P , and a set of candidate endnodes are chosen at $p=[l_{\min}, l_{\min+1}, \dots, l_{\max}]$ in P . The graph is constructed such that all neighbors of the seed are at $p=1$, all of their neighbors are at $p=2$, and in general all neighbors of node v_i at p_i lie at p_i+1 . It follows from this formulation that v_i is only reachable from the seed in p_i steps, i.e., by paths that contain p_i edges. An optimal path algorithm will find the path from the seed to the candidate endnode that results in the path of minimal cost, and because the graph is structured so that the only paths that exist between the seed and endnodes contain a bounded number of edges, it is guaranteed that the optimal path will have length on the interval $[l_{\min}, l_{\max}]$. This concept can be extended to apply length constraints on each individual locality. This ensures that the lengths of localities in the optimal path, and the length of the path itself, are within specified bounds. The bounds on the length of each locality are chosen as the range of the lengths of each locality in the training set.

The cost function includes model-based terms but can also include other model-independent, image-based terms, such as vesselness filters (see e.g. Frangi et. al. [6]). In this work, a directed vesselness filter is used, where the filter is aligned with the expected curve orientation given by the model. The analytical form of the segmentation cost function associated with edge e connecting node v_i to v_j is:

$$C(e_{i,j}) = \sum_{l=0}^{L-1} \omega_{s_j,p_j}^l \frac{g^l}{\sigma_{s_j,p_j}^l} + \sum_{k=0}^{K-1} \omega_{s_j,p_j}^{k+L} \frac{(x^k - f_{s_j,p_j}^k)^2}{(\sigma_{s_j,p_j}^k)^2}. \quad (1)$$

The second term is the cost associated with the K model-based features, while the first is associated with the L image-based features. f_{s_j,p_j}^k and σ_{s_j,p_j}^k are the mean and standard deviation of the k th model-based feature associated with sample s_j and locality p_j . x^k is the value of that feature evaluated in the target image. σ_{s_j,p_j}^l is the standard deviation of the l th image-based feature, and g^l is the value of that feature in the target image. The ω 's are the weighting values associated with each of those

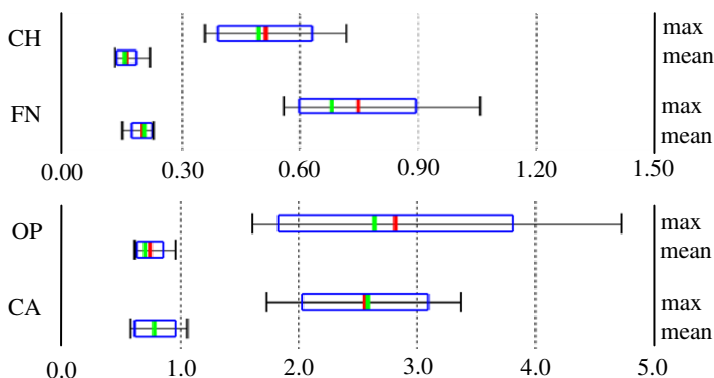


Fig. 2. Quantitative segmentation results in mm shown for the (top graph) CH, FN, (bottom graph) OP, and CA

features. The standard deviations are used as normalizing factors to bring all the features into roughly the same scale. The standard deviations and weights for the image-based features are computed with the same approach used for the model-based features, as described in the previous section. With the graph and cost function defined, the algorithm can be used to segment new images.

3 Results

To validate the approach, the algorithm was tested on four structures: the facial nerve (FN), chorda tympani (CH), and carotid artery (CA) in CT, and the optic pathway (OP) in MR/CT. All of these structures pose a significant problem for segmentation algorithms because they have intensity characteristics that vary along their length and borders that lack contrast. The FN and CH are tiny nerves in the ear that range from 5 to 1 voxel in width. The structures lie in pneumatized bone, and surrounding structures have highly variable topology. The CA is large in comparison, and is visible in CT in the temporal bone region. The OPs are composed of the optic nerves, chiasm, and tracts. This structure suffers from lack of contrast in both MR and CT. In this study, one OP is defined as one optic nerve connecting to the contra-lateral tract through the chiasm. Previous studies have shown that localizing both OPs in this fashion and computing the union of the two can lead to an accurate segmentation of the whole optic system [7]. The FN, CH, CA, and OP are divided into 6, 4, 5, and 7 localities, which roughly corresponds to the number of visually distinct regions of the structures. For each SOI, a training set of 7 binary segmentations was acquired for training, and a leave-one-out approach was used for validation. Voxel resolution for the FN and CH images was $0.3 \times 0.3 \times 0.4 \text{ mm}^3$ and was $1.0 \times 1.0 \times 1.5 \text{ mm}^3$ for the CA and OP images. Quantitative error between manual and algorithmic segmentations was measured using symmetric surface distances as shown in mm in Figure 2. In the figure, mean and max surface errors are shown for each structure. The green bar, red

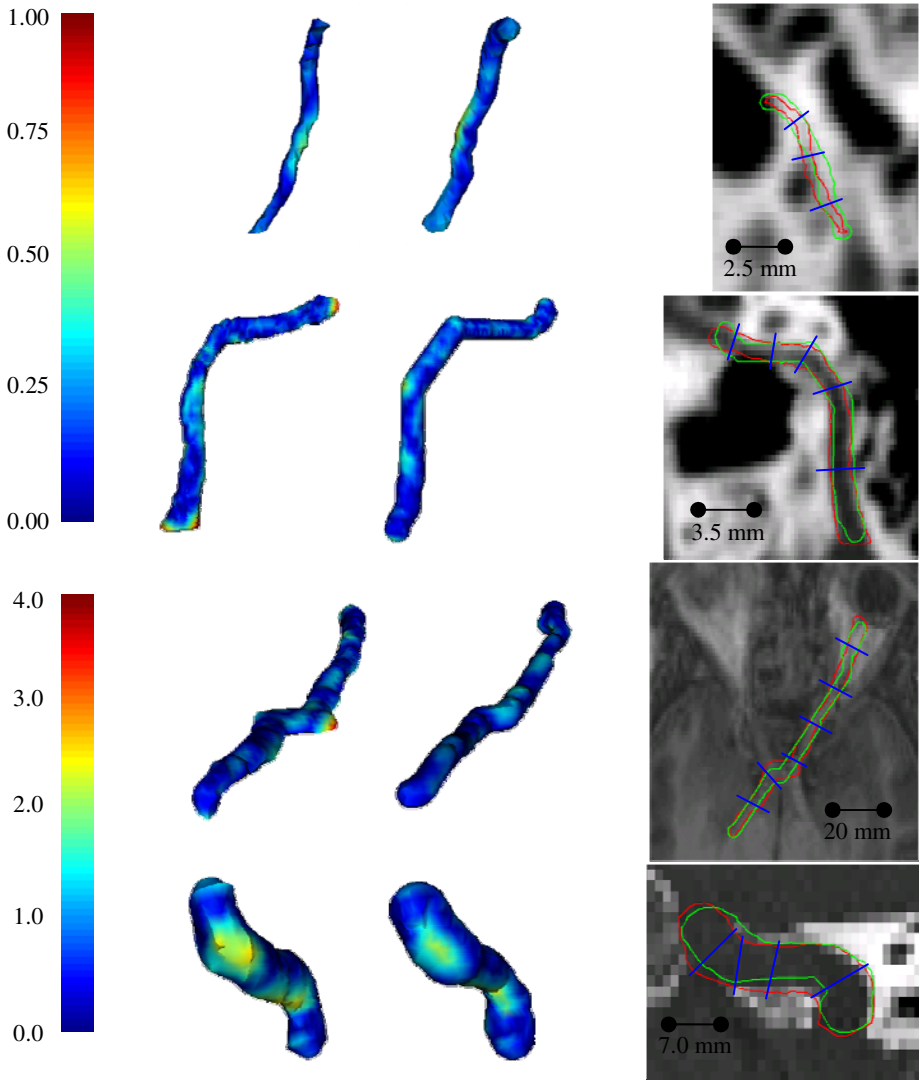


Fig. 3. Renderings of the test case resulting in worst maximum errors for the (top-to-bottom) CH, FN, OP, and CA. Left and middle columns are 3D renderings of the manual and algorithmic segmentations, color-encoded with error distance in mm as indicated by the color-bars. The right column shows contours of the manual (red) and algorithmic (green) segmentations and the automatically generated locality divisions (blue) in the images.

bar, blue rectangle, and black I-bar indicate the median, mean, one standard deviation from the mean, and the range of the errors, respectively. Each dotted-line division in the graph indicates an error of approximately 1 voxel. As seen in the graph, mean errors are sub-voxel, and the mean maximum errors are under 3 voxels for all SOIs.

Qualitative results are shown in Figure 3 for the test case that results in the worst quantitative error. To generate the 2D views, a thin-plate spline transformation was computed that warps the medial axes of the structures to lie in a plane, and then the image data and contours were passed through this transformation. This was done so that a cross section of the entire structure could be viewed in one 2D plane. As the example demonstrates, the primary location of error maxima for the FN occurs at the endpoints of the structure. This is due to errors in endpoint selection. Error maxima for the CA occur at the sharp corners of the structure, where, even though the path length is bounded, the path still may tend to try to shortcut in these regions within the allowable length bounds. Error maxima for the OP occurs at the corners of the chiasm. In this region some shortcutting occurs, and the anatomy is more rectangular in shape, which makes the tubular segmentation model less accurate. The worst errors of about 4 voxels, as demonstrated by the quantitative results and by the examples in Figure 3, are rare and highly localized. On average, results appear to be qualitatively accurate, and excellent agreement is seen between the algorithmic and manually generated contours. To demonstrate the performance of the automatic locality extraction process used to build the model, the locality divisions of the manual segmentations are also shown in the figure. As seen in the figure, the locations of the locality divisions generally correspond to changes in local intensity or curve orientation.

4 Conclusions

In this work, a novel approach for tubular structure segmentation was presented. An optimal path algorithm is used to identify the SOI in a unique high-dimensional graph that incorporates information from a novel anatomy specific tubular structure model. The graph and structure model are used to overcome some limitations inherent to approaches typically used when applying an optimal path finding algorithm. Similarly to the work of Li and Yezzi [3], the graphs presented herein allow a 3D representation of the surface to be recovered from the optimal curve. These graphs also allow local, a priori, intensity and shape information from SOI exemplars to be included and optimized by the graph search. Finally, using a priori knowledge of the approximate length of the SOI, the graphs bound the allowable structure length to remove the bias towards straight curves inherent to traditional methods. While mild shortcutting may still occur, total failure is avoided.

An approach for training the novel structure model was also presented. Using a procedure to find correspondence across exemplars, detect unique structure subsections, and extract and train discriminant features, a model of the SOI is constructed that contains statistical values and importance weightings for a set of intensity and shape features for each subsection of each training exemplar. The information in this model is then incorporated into the segmentation framework. Combining the advantages achieved by the graph structure with the weighting of the features from the model used in the cost function accounts for all four limitations of traditional graph-based approaches listed in Section 1.

For validation, the approach was tested on four challenging anatomical structures. As can be seen in the images of Figure 3, there is poor image contrast at the borders

of these structures and changing intensity profiles along the lengths of the structures. There are also substantial inter-patient variations in the intensity and shape features for each of these structures. All of these factors make the segmentation problem very difficult. However, the results of the presented approach are accurate, suggesting that it is generally applicable for fixed topology, non-branching tubular structures. To the extent of our knowledge, there are only a few existing methods for the segmentation of the FN, CH, and OP. The segmentation results presented in this paper are comparable to those of the most accurate of those existing approaches, which were specifically designed to localize these structures [5,7]. This also indicates that the approach presented in this paper is effective.

The method does require selecting one single parameter: the number of localities of the SOI. However, its selection is intuitive, i.e., it should roughly correspond to the number of visually distinct regions of the SOI. In future work, this will be chosen automatically. One possibility is brute force optimization, i.e., try every possible number of localities between 1 and N on the training set, and use the one that produces the most accurate results.

Acknowledgements. This research has been supported by NIH grants R01EB006193 and F31DC009791. The content is solely the responsibility of the authors and does not necessarily represent the official views of these institutes.

References

1. Lesage, D., Angelini, E., Bloch, I., Funka-Lea, G.: A review of 3D vessel lumen segmentation techniques: Models, features, and extraction schemes. *Medical Image Analysis* 13, 819–845 (2009)
2. de Bruijne, M., van Ginneken, B., Viergever, M.A., Niessen, W.J.: Adapting active shape models for 3D segmentation of tubular structures in medical images. In: *Proc. Inf. Process. Med. Imaging*, pp. 136–147. Springer, Heidelberg (2003)
3. Dijkstra, E.W.: A note on two problems in connexion with graphs. *Numerische Mathematik* 1, 269–271 (1959)
4. Li, H., Yezzi, A.: Vessels as 4D Curves: Global Minimal 4D Paths to Extract 3D Tubular Surfaces and Centerlines. *IEEE Trans. Med. Imag.* 26(9), 1213–1223 (2007)
5. Noble, J.H., Warren, F.M., Labadie, R.F., Dawant, B.M.: Automatic segmentation of the facial nerve and chorda tympani in CT images using spatially dependent feature values. *Medical Physics* 35(12), 5375–5384 (2008)
6. Frangi, A.F., Niessen, W.J., Vincken, K.L., Viergever, M.A.: Multiscale vessel enhancement filtering. In: Wells, W.M., Colchester, A., Delp, S.L. (eds.) *MICCAI 1998*. LNCS, vol. 1496, pp. 130–137. Springer, Heidelberg (1998)
7. Noble, J.H., Dawant, B.M.: An atlas-navigated optimal medial axis and deformable model algorithm (NOMAD) for the segmentation of the optic nerves and chiasm in MR and CT images. *Medical Image Analysis* (2011, in press), doi:10.1016/j.media.2011.05.001

Learning-Based Meta-Algorithm for MRI Brain Extraction

Feng Shi¹, Li Wang¹, John H. Gilmore², Weili Lin³, and Dinggang Shen¹

¹IDEA Lab, ³MRI Lab, Department of Radiology and BRIC,
University of North Carolina at Chapel Hill, USA

²Department of Psychiatry, University of North Carolina at Chapel Hill, USA
dgshen@med.unc.edu

Abstract. Multiple-segmentation-and-fusion method has been widely used for brain extraction, tissue segmentation, and region of interest (ROI) localization. However, such studies are hindered in practice by their computational complexity, mainly coming from the steps of template selection and template-to-subject nonlinear registration. In this study, we address these two issues and propose a novel learning-based meta-algorithm for MRI brain extraction. Specifically, we first use exemplars to represent the entire template library, and assign the most similar exemplar to the test subject. Second, a meta-algorithm combining two existing brain extraction algorithms (BET and BSE) is proposed to conduct multiple extractions directly on test subject. Effective parameter settings for the meta-algorithm are learned from the training data and propagated to subject through exemplars. We further develop a level-set based fusion method to combine multiple candidate extractions together with a closed smooth surface, for obtaining the final result. Experimental results show that, with only a small portion of subjects for training, the proposed method is able to produce more accurate and robust brain extraction results, at Jaccard Index of 0.956 ± 0.010 on total 340 subjects under 6-fold cross validation, compared to those by the BET and BSE even using their best parameter combinations.

1 Introduction

Brain extraction, also called as skull stripping, is a fundamental step in the brain MR image preprocessing, aiming to remove the non-brain tissues, such as skull, scalp, and dura. Accurate brain extraction is crucial since it is non-reversible for the subsequent steps to recover the wrongly-removed brain tissues, and also the residual non-brain tissue (especially dura) in the brain extraction result could overestimate the local brain volume or cortical thickness. Accordingly, a number of automated brain extraction algorithms have been developed, using morphology, morphology combined with edge detection (BSE) [1], deformable model (BET) [2], graph cut, watershed, and other hybrid techniques. Each algorithm has its merits and pitfalls. For example, it is reported that BET usually remains some non-brain tissue, while BSE may remove extra brain tissue [3]. Thus, when applied to a large cohort with varying scanning parameters and diagnosis types, manual adjustment of program parameters and manual editing of extraction results are inevitable.

Multiple-segmentation-and-fusion method has been recently developed with promising results in brain extraction [4], tissue segmentation [5], and region of interest (ROI) localization such as hippocampus [6]. Generally, the method choose a subset of the most similar templates from the template library to propagate their labels to the test subject, and then fuse them together for a final labeling. Studies suggest that combining complementary results may substantially improve the robustness of final brain extraction. However, the applications of this method are largely hindered by the high computational complexity, mainly coming from the first two steps, namely template selection and template-to-subject propagation (usually by nonlinear registration). For example, in MAPS method [4], it takes about 19 hours for brain extraction of a single subject.

In this study, we address these issues by proposing a novel learning-based meta-algorithm for MRI brain extraction. Specifically, first, we do not need exhaustive template-subject comparison to select the appropriate templates; instead, we use exemplars to represent the entire template library, and thus the test subject just needs to be compared with exemplars. Second, without need of the template-to-subject warping, we use a meta-algorithm, i.e., combining two existing individual brain extraction algorithms (BET and BSE), to conduct multiple extractions directly on test data. Effective parameter settings for the meta-algorithm are learned from the training data and passed to subject through exemplars. We further develop a level-set based label fusion to combine multiple candidate extractions together with a closed smooth surface, for obtaining the final result. Thus, errors in voxel-wise label fusion (e.g., by simple majority voting) can be greatly eliminated. Methodological details and experimental results are provided in the sections below.

2 Materials and Method

2.1 Subjects and MRI Acquisition

Our data is obtained from Alzheimer's Disease Neuroimaging Initiative (ADNI) (www.adni-info.org), which is a large multi-site study acquired from multiple centers, and so it is ideal to test the accuracy and robustness of the proposed brain extraction algorithm. We use the baseline T1-weighted MRI scans of 340 subjects with age ranging from 55 to 90 years old, including 33 Alzheimer's disease (AD), 236 mild cognitive impairment (MCI), and 71 healthy controls. These MRI scans were acquired from 1.5T scanners made by GE, Philips, and Siemens, using the scanning parameters: TR/TI/TE=2300/1000/3.5 ms, flip angle=8°, FOV=240×240 mm². Two sizes of images were acquired: one is 160 sagittal slices with voxel resolution of 1.25×1.25×1.2 mm³, and another is 180 sagittal slices with voxel resolution of 0.94×0.94×1.2 mm³.

For preprocessing, all images are resampled into isotropic 1×1×1 mm³. Bias field is estimated with N3 algorithm [7] to correct the intensity inhomogeneity in all MR images separately.

2.2 Overview of the Proposed Learning-Based Meta-Algorithm

A learning-based meta-algorithm is proposed for brain extraction, with three major steps as illustrated in Fig. 1. A small number of subjects are selected as training data, and the rest are used as test data. The *first* step is the brain localization, where a brain mask is generated from the training data to roughly localize the brain region of the test subject. The *second* step is exemplar selection and parameter learning. Exemplars are chosen for representing the whole training data. Parameters for each individual algorithm (BET/BSE) are uniformly sampled and exhaustively combined, to learn the parameter map of brain extraction performance for each training subject. As for each exemplar, its parameter map can be obtained by averaging from all training subjects according to their contributions to that exemplar. The *third* step is the brain extraction and fusion for the test subject. In the application, each test subject is compared with all exemplars to find the best matching one, and then the corresponding set of parameter combinations are employed to conduct multiple brain extractions on the test subject. All candidate extractions are then fused together into the final result with a level-set based algorithm. Details are given in the following subsections.

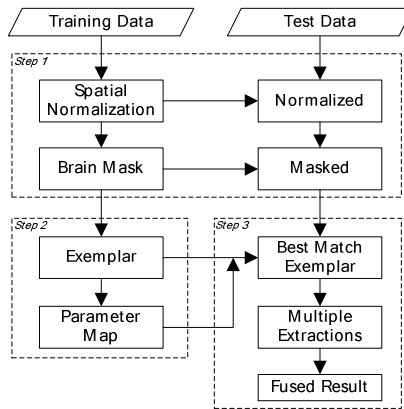


Fig. 1. A conceptual diagram illustrating the three main steps in the proposed learning-based meta-algorithm for brain extraction

2.2.1 Step 1: Brain Localization

This step is to spatially normalize the training data and generate a mask to roughly localize the brain for the test subject, so that the search range can be reduced and individual brain extraction algorithms can perform better. For example, when applied to the original brain image, BET often keep some non-brain tissues in front of brain stem, and BSE may even remain some neck regions [8]. These problems can be partially solved by this step.

Each training subject contains two images, namely a with-skull image and a brain-extracted image, which is manually delineated and serves as the ground-truth. First, all the brain-extracted images of training subjects are affine aligned onto a widely used population template named as ICBM152 [9] for spatial normalization. The estimated transformations are used to bring their with-skull images also onto the common

space. Then, the with-skull images and the brain-extracted images are averaged across all subjects separately. The average brain-extracted image (Fig. 2a) is binarized into a brain mask by keeping all non-zero voxels, which is further dilated for a number of voxels d to generate a final brain mask for better brain localization (Fig. 2b). We conservatively set $d = 12$ in this paper to avoid removing brain tissues when normalization errors exist. The average with-skull image (Fig. 2c) is used to serve as an intensity template for aligning the test image (which only has the with-skull image) onto the same space (Fig. 2d). Finally, the warped test image is masked with the generated brain mask (Fig. 2e). By doing this, the non-brain tissues with a certain distance from the brain in the test image can be completely (e.g., neck regions) or partially (e.g., near the brain stem regions) removed.

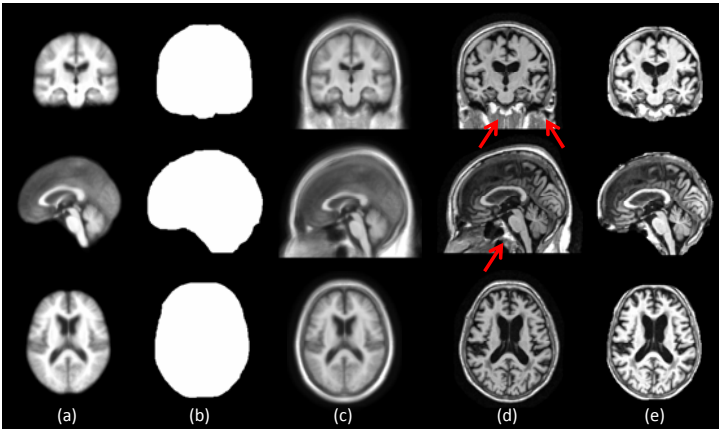


Fig. 2. Illustration of the average brain-extracted image from training data (a), the generated brain mask (b), the average with-skull image from training data (c), a test image from an AD subject (ID 067_S_0110) after affine alignment (d), and the test image applied with brain mask (e). Red arrows indicate some non-brain tissues, which pose challenges to individual brain extraction algorithms, are totally or partially removed by the brain mask.

2.2.2 Step 2: Exemplar Selection and Parameter Learning

Exemplar Selection. We propose to use a small number of exemplars to represent the entire data, so that prior knowledge can be carried by these exemplars and then propagate to the test subject. By doing so, the traditional exhaustive template-subject comparisons can be substantially reduced. To do this, we mask the with-skull images of training data by the dilated brain mask. Then, we use the affinity propagation algorithm [10] to cluster the masked training data into a number of classes, and further determine their respective exemplars. Briefly, each pair of training data is computed for their intensity similarity using mutual information, and the results on all possible pairs form an N by N matrix, where N is the number of training data. The diagonal elements of the similarity matrix have influence on the number of selected exemplars, which are set as the median of the input similarities to allow all training data to have equal probability as exemplars. In each iteration, the elements in similarity matrix

exchange messages of how suitable one element to be the exemplar of another one. When the matrix updating converges, a number of training subjects are selected as exemplars.

Parameter Learning. We propagate only the parameter settings learned from the training data to conduct multiple extractions directly on the test subject. BET and BSE are employed in our meta-algorithm due to their well-recognized performance and speed. Each of them has a number of parameters. We uniformly sample the commonly-used parameters for the two algorithms, as specified in Table 1. Then, every combination of parameters is used to conduct brain extraction on each masked training data. The overlap rate is finally computed between automated segmentation A and manual segmentation M with Jaccard Index as $J = |A \cap M|/|A \cup M|$. Thus a parameter map is formed for each individual algorithm on each training data, e.g., a 15×6 map built for BET (Fig. 3a) and a $5 \times 4 \times 5$ map built for BSE, where each element represents the overlap rate with respect to the parameters used. For each exemplar, its parameter map (Fig. 3b) is defined as the average over all training data according to their contributions to that exemplar.

Table 1. Parameters and sampling strategies used for BET and BSE in the proposed meta-algorithm

Methods	Parameters	Default	Sampling
BET	-f (fractional intensity threshold)	0.5	0.1:0.05:0.8
	-g (vertical gradient)	0	-0.3:0.1:0.2
BSE	-d (diffusion constant)	25	15:5:35
	-n (diffusion iterations)	3	1:2:7
	-s (edge detection constant)	0.62	0.5:0.05:0.7

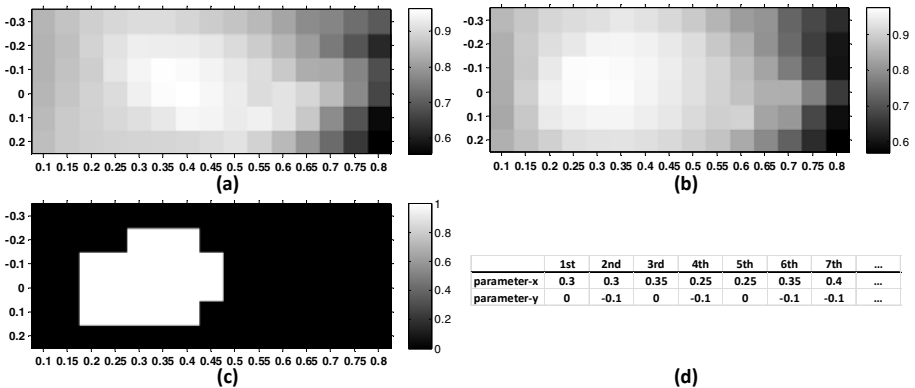


Fig. 3. Illustration of BET parameter maps on a training data (a) and an exemplar (b), as well as the thresholded exemplar parameter map showing top M highest ranking parameter combinations (c-d). X-axis and y-axis represent the two parameters (i.e., -f and -g) in BET.

2.2.3 Step 3: Brain Extraction and Fusion

Brain Extraction. For the test subject, the best match exemplar is found by comparing the intensity similarity of its masked with-skull image with that of all exemplars. Thus, the parameter map of the selected exemplar can be used to guide multiple brain extractions on the test subject. Top M effective parameter combinations are chosen from the parameter map (Fig. 3c), so that M instances of brain extractions are conducted on the test subject (Fig. 3d). Note that more instances of individual brain extractions will bring higher computational cost, although the accuracy might be improved. We empirically use $M = 20$ in this paper.

Fusion. When multiple candidate extractions are available, the next question is how to fuse them together. Majority voting is widely used; however, it can produce isolated false extractions and sometimes unsmooth boundary in the extracted brain region. To address this issue, we model the brain as a closed entity with a smooth surface, and thus use a level-set method [11] to find the brain boundaries from the average label map of the multiple brain extraction results, as shown in Fig. 4. For this method, two parameters are important in constraining the shape of the level-set based surface. The first parameter is to smooth the resulting level set with kernel s , and the second parameter is to dilate the final brain extraction result outward for t voxels to reduce the risk of cutting brain tissues, as recommend in many studies [4]. It is worth noting that these two parameters could also be learned from the training data, by uniformly sampling the parameter values and picking the one with the best overlap with ground truth. We use $s = 11$ and $t = 2$ in this paper. Finally, by applying the estimated level-set mask (Fig. 4d) onto the test subject, the brain extraction result can be obtained (Fig. 4e).

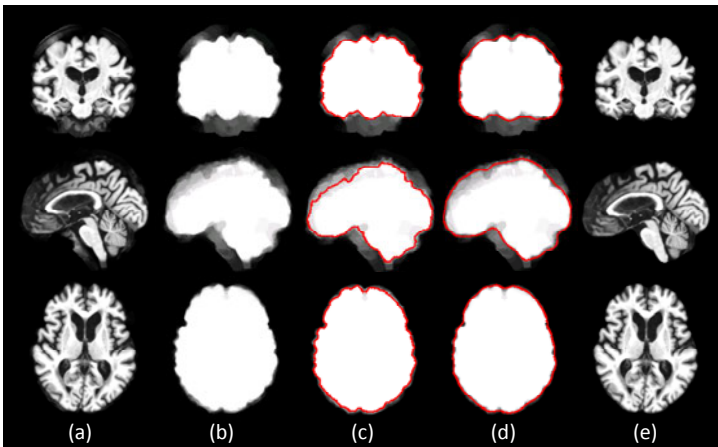


Fig. 4. Illustration of the level-set based fusion on the AD subject shown in Fig. 2. Averaged candidate extractions is shown in (a), averaged mask in (b), level-set based surfaces (red) using $s = 0$, $t = 0$ (c), or $s = 11$, $t = 2$ (d), and final extraction result (e).

3 Experimental Results

In our experiment, a 6-fold cross-validation is performed. All subjects were randomly partitioned into 6 folds. Each fold was used in the training process and the rest was left as testing set. Estimation of errors was the overall proportion of errors committed in all folds. Taking one case as example, 54 of 340 subjects (8 AD, 22 MCI, and 24 controls) are selected as training data, and the rest are used as test data. With the method described above, the exemplars we obtained come from 2 AD, 3 MCI, and 1 control, indicating in part the ability of our exemplars for representing the data distribution.

Automated extractions are compared with manual ones by calculating their overlap rates. The performance is further compared with that obtained by each of the four baseline methods, i.e., BET without/with our brain localization step (called Raw BET and BET, respectively), and also BSE without/with our brain localization step (called Raw BSE and BSE, respectively). For these four baseline methods, every combination of parameters is explored and the one with the highest average Dice ratio across the entire test data is chosen as the final result. Furthermore, majority voting is also used to replace the level-set based data fusion in the proposed pipeline, and serve as another control method. All results are shown in Fig. 5. It can be observed that, the proposed method significantly outperforms other five methods ($p < 0.001$) by achieving the average Jaccard Index of 0.956 ± 0.010 , in comparison with 0.903 ± 0.026 and 0.935 ± 0.010 by Raw BET and BET, 0.870 ± 0.029 and 0.883 ± 0.037 by Raw BSE and BSE, and 0.849 ± 0.036 by MajorityVote. Note that, by using our method, not only the accuracy is largely improved, the robustness is also increased as reflected by the lowest standard deviation. Meanwhile, our brain localization step shows its ability to improve the accuracy and robustness of brain extractions for both BET and BSE, compared with Raw BET and Raw BSE, respectively. Another observation is that the extraction result on abnormal brain such as AD has lower accuracy than that of control, which agrees with previous findings [3].

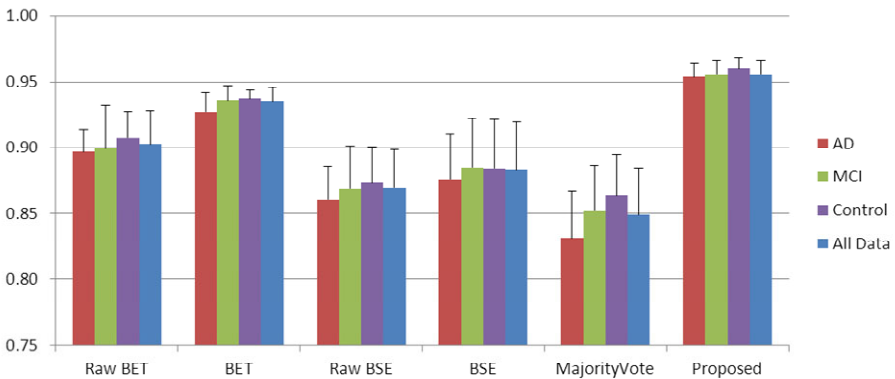


Fig. 5. Jaccard Index for automated brain extraction by the four baseline methods, majority voting, and the proposed method, under 6-fold cross validation

To visualize the location of extraction errors in 3 algorithms (BET with brain localization, BSE with brain localization, and the proposed method), we generate the projection maps for the false positive and negative voxels. Specifically, the 3D false positive (or negative) maps of all test data are first averaged, and then projected onto the sagittal, coronal, and axial directions, as shown in Fig. 6. BET has more false positive voxels, indicating the inclusion of more non-brain tissues in the brain extraction results. BSE has more false negative voxels, indicating wrongly removing more brain tissues. Our proposed method produces the lowest false positive and false negative, in comparison to both baseline algorithms.

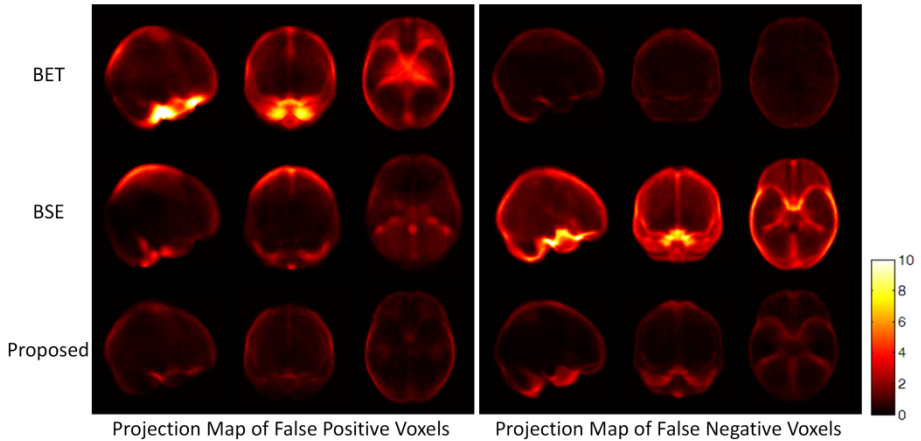


Fig. 6. Projection maps of false positive (left panel) and false negative (right panel) for the brain extractions obtained by BET, BSE, and our proposed method on all test data. Brighter value indicates more false positive (or negative) results in that location.

4 Discussion and Conclusion

We have presented a novel learning-based meta-algorithm for MRI brain extraction. It has three novelties. First, we introduce exemplars to represent the whole template library, thus largely reducing the size of templates to match during the application stage. Second, we alternatively propagate the learned parameter settings from template to subject, without requiring nonlinear registration between them. Third, we develop a level-set based label fusion method to overcome both boundary discontinuity and isolated errors in voxel-wise label fusion methods. Experimental results have shown that our proposed method can substantially improve the accuracy and robustness of brain extraction. Also, the proposed meta-algorithm is not limited to BET and BSE; it can be extended for inclusion of other brain extraction algorithms.

The proposed method has demonstrated comparable brain extraction accuracy (Jaccard 0.956, tested on 340 subjects) with MAPS (Jaccard 0.954, tested on 839 subjects) [4]. More importantly, it has much shorter processing time (~ 10 mins) to perform brain extraction on a new subject than that of MAPS (~ 19 hours) [4]. For training stage, 6 hours are needed to process 54 training data on a standard PC for spatial normalization, parameter map learning, and exemplar selection.

Acknowledgments. This work was supported in part by NIH grants EB006733, EB008760, EB008374, EB009634, MH088520, MH070890, MH064065, NS055754, and HD053000.

References

1. Shattuck, D.W., Leahy, R.M.: Automated graph-based analysis and correction of cortical volume topology. *IEEE Trans. Med. Imaging* 20(11), 1167–1177 (2001)
2. Smith, S.M.: Fast robust automated brain extraction. *Hum. Brain Mapp.* 17(3), 143–155 (2002)
3. Fennema-Notestine, C., Ozyurt, I.B., Clark, C.P., Morris, S., Bischoff-Grethe, A., Bondi, M.W., Jernigan, T.L., Fischl, B., Segonne, F., Shattuck, D.W.: Quantitative evaluation of automated skull-stripping methods applied to contemporary and legacy images: effects of diagnosis, bias correction, and slice location. *Hum. Brain Mapp.* 27(2), 99–113 (2006)
4. Leung, K.K., Barnes, J., Modat, M., Ridgway, G.R., Bartlett, J.W., Fox, N.C., Ourselin, S.: Brain MAPS: An automated, accurate and robust brain extraction technique using a template library. *NeuroImage* 55(3), 1091–1108 (2010)
5. Weisenfeld, N.I., Warfield, S.K.: Automatic segmentation of newborn brain MRI. *Neuroimage* 47(2), 564–572 (2009)
6. Leung, K.K., Barnes, J., Ridgway, G.R., Bartlett, J.W., Clarkson, M.J., Macdonald, K., Schuff, N., Fox, N.C., Ourselin, S.: Automated cross-sectional and longitudinal hippocampal volume measurement in mild cognitive impairment and Alzheimer’s disease. *NeuroImage* 51(4), 1345–1359 (2010)
7. Sled, J.G., Zijdenbos, A.P., Evans, A.C.: A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Trans. Med. Imaging* 17(1), 87–97 (1998)
8. Rex, D.E., Shattuck, D.W., Woods, R.P., Narr, K.L., Luders, E., Rehm, K., Stolzner, S.E., Rottenberg, D.A., Toga, A.W.: A meta-algorithm for brain extraction in MRI. *NeuroImage* 23(2), 625–637 (2004)
9. Mazziotta, J.C., Toga, A.W., Evans, A., Fox, P., Lancaster, J.: A probabilistic atlas of the human brain: theory and rationale for its development. *Neuroimage* 2(2), 89–101 (1995)
10. Frey, B.J., Dueck, D.: Clustering by passing messages between data points. *Science* 315(5814), 972–976 (2007)
11. Chan, T.F., Vese, L.A.: Active contours without edges. *IEEE Trans. Image Process.* 10(2), 266–277 (2001)

Learning Likelihoods for Labeling (L3): A General Multi-*Classifier* Segmentation Algorithm

Neil I. Weisenfeld and Simon K. Warfield

Computational Radiology Laboratory,
Children's Hospital Boston,
Harvard Medical School,
Boston, MA

Abstract. PURPOSE: To develop an MRI segmentation method for brain tissues, regions, and substructures that yields improved classification accuracy. Current brain segmentation strategies include two complementary strategies. Multi-spectral classification techniques generate excellent segmentations for tissues with clear intensity contrast, but fail to identify structures defined largely by location, such as lobar parcellations and certain subcortical structures. Conversely, multi-template label fusion methods are excellent for structures defined largely by location, but perform poorly when segmenting structures that cannot be accurately identified through a consensus of registered templates. METHODS: We propose here a novel multi-*classifier* fusion algorithm with the advantages of both types of segmentation strategy. We illustrate and validate this algorithm using a group of 14 expertly hand-labeled images. RESULTS: Our method generated segmentations of cortical and subcortical structures that were more similar to hand-drawn segmentations than majority vote label fusion or a recently published intensity/label fusion method. CONCLUSIONS: We have presented a novel, general segmentation algorithm with the advantages of both statistical classifiers and label fusion techniques.

1 Background

Vannier introduced multi-spectral classification for brain image segmentation in 1985 [1] and such segmentation strategies have been the standard for years (e.g. [2,3]). Multi-spectral classification yields segmentations with true fidelity with respect to the image data and the patient, but are unable to label structures where the boundaries are poorly represented by changes in image contrast, such as lobar parcellations and subcortical substructures. With advances in flexible, non-rigid registration technology, multi-atlas approaches have recently become popular (e.g. [4,5,6]). These are segmentation-by-registration approaches where registration error is marginalized by fusion of a series of registered template segmentations through some consensus generating process. These methods ultimately have limited ability to resolve structure that cannot be identified through

registration alone. Recent work [6] has attempted to address this shortcoming by adding a grayscale image similarity term to the label fusion process. This technique is limited by the simplifying assumption that template image intensities will be similar to subject image intensities, and do not account for the significant differences in image intensity between scanners, pulse sequences, and disease states.

We propose a new method called Learning Likelihoods for Labeling (L3), which unifies label fusion and statistical classification. In our approach, each input template does not define a candidate segmentation to be fused, but instead defines a unique *classifier* that is used to generate a multi-spectral classification of the target subject. Each of the resulting classifications is then fused, yielding an individualized segmentation true to the underlying data. Our strategy is capable of labeling both types of structures: tissues that are well-defined by the contrast in the images, as well as labels defined by their relative location to other structures. We present below a self-contained learning-by-example strategy for segmentation where the only inputs are a small number of segmented and registered example images and the target MRI to be segmented. We demonstrate this approach on data from a group of 14 subjects for whom gray matter, white matter, ventricles, putamen, caudate, and thalamus have been hand-labeled by an expert.

2 Methodology

2.1 Overview

Our algorithm utilizes a library of template labelmaps $\{L_n\}, n = 1, \dots, N$ which have been registered to our target subject. In the present work, we employ a particular non-linear, non-rigid registration approach [7] for aligning the individual library images to the target, but a number of excellent techniques could have been used.

A typical label fusion approach seeks to combine these labels to generate a consensus labelmap or estimated true segmentation $T = f(L_1, \dots, L_n)$ using some fusion function f , such as majority voting or STAPLE [8]. This approach is limited, however, in its use of intensity information in the images and is therefore only able to segment features well-represented in the template images. By using the template images to guide training of a supervised classifier, we're able to "learn" about patterns of intensity throughout the image. We also utilize each template individually as a spatial prior, and in so doing retain the benefits of a traditional label fusion strategy where anatomical variation is represented by the distribution of input templates.

We assume that we have aligned, multi-spectral image data from the subject we wish to segment and denote this vector image \mathbf{I} . We assume the existence of a supervised classification algorithm $C : L \times \mathbf{I} \rightarrow S$ which takes a labelmap for training L , a multi-spectral dataset \mathbf{I} and produces a segmentation S . Coupling C with a particular template labelmap L_n yields a new classifier $C_n : \mathbf{I} \rightarrow S_n$ that generates a candidate segmentation S_n . We fuse these *classifications* to produce the final resulting segmentation $T = f(S_1, \dots, S_n)$.

2.2 Classification

In order to generate classifications $\{S_n\}$, we implement a Bayesian segmentation strategy:

$$p(S_{ni} = s | \mathbf{I}_i) = \frac{p(\mathbf{I}_i | s)p(s)}{\sum_{s'} p(\mathbf{I}_i | s')p(s')} \quad (1)$$

and estimate the likelihood $p(\mathbf{I}_i | s)$ of a particular MR intensity value \mathbf{I}_i , at voxel i , given a particular tissue class s , using the density estimation strategy due to Cover [9]. Training data for density estimation is generated by sampling each labelmap L_n at a random coordinate i and pairing this label L_{ni} with the target subject's image intensity at the same coordinate \mathbf{I}_i . A number of such samples, randomly distributed in space, are then employed to estimate the tissue class likelihoods. In the experiments that follow, we used 4000 samples per tissue class and $k = 51$ for the Cover algorithm.

For the prior in Eq. 1, we use a spatially varying prior derived from an individual template L_n as in [6]. If D_{ni}^s is the signed distance transform of label s in input template L_n , at voxel i , then the prior $p(s) = \frac{1}{Z} \exp(\rho D_{ni}^s)$ where Z is the appropriate normalization constant such that all probabilities sum to 1 and ρ is a parameter which allows us to control the smoothness of the prior. The ideal value of ρ is related to the expected registration error, and for the non-linear registration we used in this work, we chose $\rho = -1.4$.

Each template labelmap L_n generates a unique segmentation S_n of the target subject that is the product of both a unique prior and likelihood. An alternative strategy might employ a common prior across candidate segmentations S_n , for instance based on the prevalence of a given label across templates. We believe that using each template individually allows us to better marginalize registration error during the fusion process.

2.3 Weighted Fusion

The STAPLE algorithm [8] has previously been used successfully for label fusion in contexts where it is not known how informative each template labelmap is on a label-by-label basis. Indexing voxels by i and individual templates using n , the voxelwise formula for fusion from STAPLE is:

$$p(T_i = s | \mathbf{S}_i, \Theta) = \frac{p(T_i = s) \prod_n p(S_{ni} | T_i = s, \theta_n)}{\sum_{s'} p(T_i = s') \prod_n p(S_{ni} | T_i = s', \theta_n)} \quad (2)$$

where θ_n is a matrix of performance parameters indicating the probability of mismatch between the individual segmentation S_n and the true, underlying segmentation T for each possible combination of labels s', s : $\theta_{ns's} \equiv p(S_n = s' | T = s)$, Θ is the collection of $\theta_1, \dots, \theta_N$, and \mathbf{S} is the collection of S_1, \dots, S_N . This equation is solved using the STAPLE algorithm [8].

The performance parameters Θ in STAPLE depict statistics accumulated across the region of computation. This region may be restricted to a window around each voxel and Θ then depicts a spatially-varying measure of performance. This measure is advantageous for label-fusion applications where true performance may vary across the image, but may lead to undesirable local optima in areas of disagreement. In [10], the authors present a maximum a posteriori (MAP) estimate for Θ . Supplying a prior for Θ stabilizes these estimates in the face of missing structures and avoids undesirable local optima in areas of extreme disagreement.

In the present application, we generated a consensus from each candidate segmentation by running STAPLE MAP locally in a $5 \times 5 \times 5$ voxel window around each voxel. The prior on θ_j is a Beta distribution, as described in [10], with parameters ($\alpha = 5, \beta = 1.5$) for the diagonal elements and ($\alpha = 1.5, \beta = 5$) for the off-diagonal elements of each θ_j . We utilized a spatially varying prior in STAPLE MAP calculated as the prevalence of each label, at each voxel coordinate, in the registered template images L_n . Finally, a mean-field approximation to a Markov Random Field prior was used as described in [8].

2.4 Improving the Training Data

Given the estimated true segmentation from Equation 2, it is desirable to edit the input training templates $\{L_n\}$ by removing voxels that are inconsistent with the estimated segmentation [11]. We do this stochastically using the following procedure: for each of our templates L_n , we examine each of the previously sampled training points i , with label L_{ni} , and remove the training point from further use if $p(T_i = L_{ni} | \mathbf{S}, \Theta) < \text{rand}([0, 1])$ where $\text{rand}([0, 1])$ is a random number selected from a uniform distribution between zero and one. For instance, if a particular training point is labeled as “gray matter,” and our estimated true segmentation indicates that the probability of this label at this voxel coordinate is 0.6, then there is a $1 - 0.6 = 0.4$ chance that we will remove this point.

Once the training data is edited, we re-estimate the classifications $\{S_n\}$ and then generate a new weighted consensus using Eq 2. This process continues iteratively until the difference in successive estimates of the final segmentation is small. In [11] we demonstrated that this procedure leads to improved classifications by removing training points which appear inconsistent with the data being segmented. This algorithm serves to detect registration error between the template and the target and removes training data from the boundaries between tissue types where registration errors occur more frequently. This leads to improved classification accuracy.

3 Validation and Experiments

An expert segmented each of 14 normal controls into gray matter, white matter, ventricular cerebrospinal fluid (including 3rd, 4th, 5th, and lateral ventricles), caudate, putamen, and thalamus following an established protocol. Segmentations were performed using high-resolution (1mm isotropic) T1-weighted SPGR

images acquired at 3T on a GE Signa Imager (General Electric, Wakeusha, WI, USA). For each template, an associated grayscale image was registered to the target subject using an implementation of rigid and affine mutual-information (MI) registration and then using a locally-affine MI-based registration [7]. The resulting registration transform was then used to warp the template segmentation to the target subject. Each of the 14 datasets was segmented using the other 13 as templates in a typical “leave-one-out” fashion.

For comparison to our algorithm, we report values from segmentations obtained with two methods: an implementation of the label/intensity fusion algorithm in [6] and majority voting. Each of the algorithms used the same registered templates and, for compatibility with the other methods, we restricted our tests in these experiments to classification using the T1-weighted intensities alone, although a potential advantage of our method is its ability to perform multi-spectral classification.

Figure 2 shows a typical T1-weighted image (top left) followed by the result from our algorithm (top right), a label/intensity fusion segmentation (bottom left), and the hand drawn segmentation (bottom right) for Case 1 from our study. The label fusion result appears to routinely oversegment cortical gray

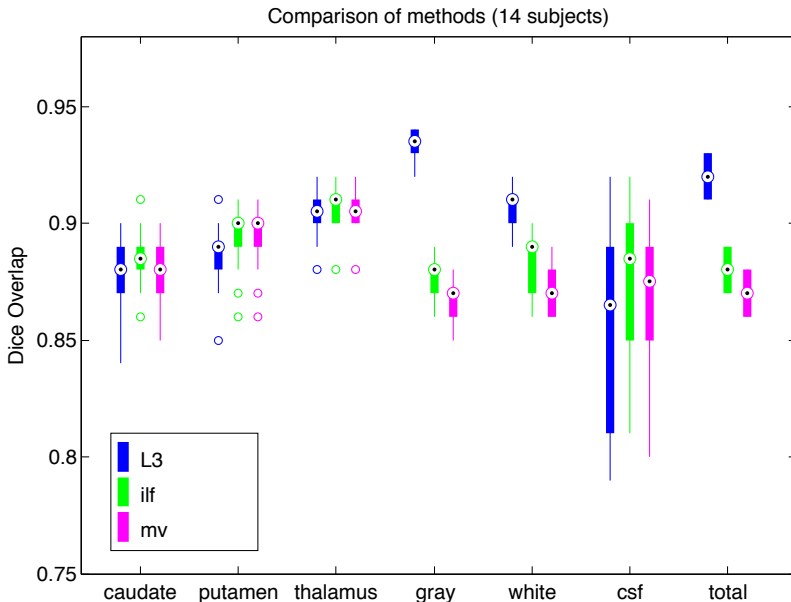


Fig. 1. Dice overlap comparison between automatic segmentations and hand-drawn images for 14 subjects. Shown are our L3 method (blue), the combined intensity/label fusion method (ilf; green), and majority voting (mv; magenta). L3 performance is similar to other methods for subcortical structures (caudate, putamen, thalamus), but shows superior performance overall due to large improvements in gray matter and white matter segmentation.

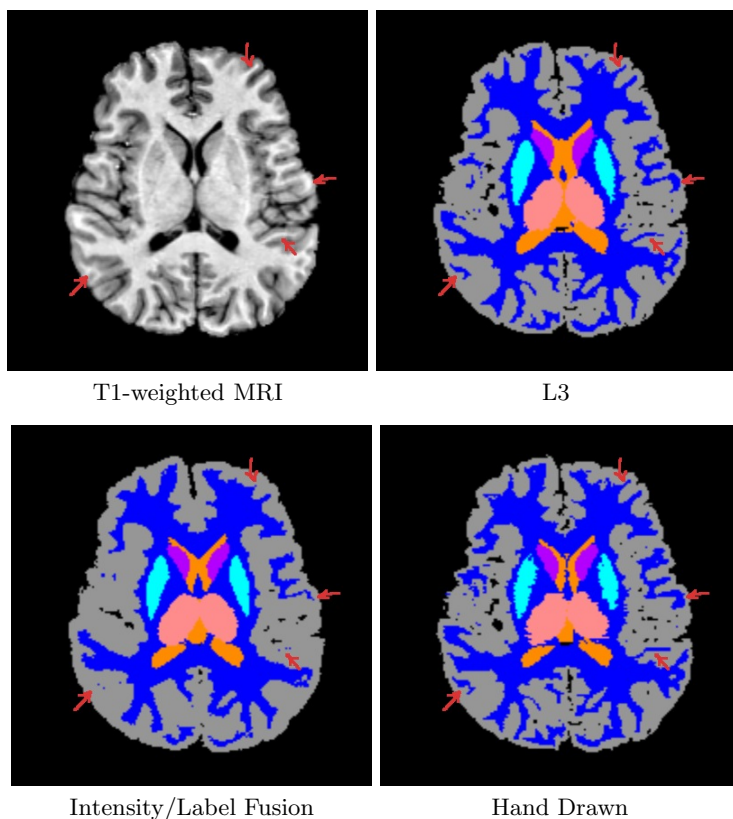


Fig. 2. Comparison of our method (L3) with a particular intensity/label fusion technique (see text) and hand-drawn segmentations. The intensity/label fusion result (lower left) oversegments gray matter and fails to represent fine sulcal detail in the image. Our method (top right) extracts such detail and better approximates the hand-drawn segmentation (bottom right).

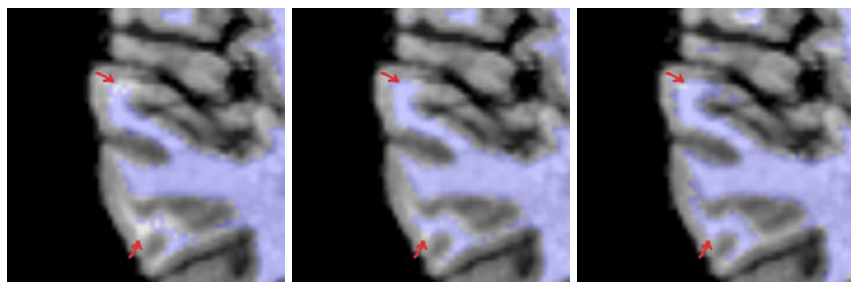


Fig. 3. The impact of editing the training sets: first iteration of our method (left), final iteration of our method (middle), groundtruth image (right). The method iteratively improves its class-conditional density estimates in order to better capture fine intensity differences within the input. Arrows indicate areas of change.

matter and misses fine details (arrows). In our algorithm, each individual input template leads to a different set of tissue-class likelihood estimates for the target subject, leading to different classifications which are then fused. We can then iteratively improve these likelihood estimates using the fused classifications. Figure 3 shows a close-up of segmentations overlaid onto the T1 image. On the left is the initial, fused classification using our method (first iteration), with arrows indicating areas where fine white matter details are not well captured. The next image shows the final result of our method, after iteratively improving the class-conditional likelihood estimates. Our new method captures more structure in the white matter as the likelihoods improve with each iteration. On the right is the hand-drawn segmentation for the corresponding region.

Figure 4 shows a plot of Dice overlap [12] measures for segmentations from our method, as well as majority voting and our implementation of the algorithm in [6], each compared with expert hand-drawn segmentations. Performance of L3 is generally excellent and is consistently highest for gray matter, white matter, and overall. The large variability in CSF segmentation is due to partial volume effects and the difficulty of visualizing CSF on T_1 weighted images. Total Dice measures across labels are computed as $\frac{\sum_s 2|A_s \cap B_s|}{\sum_s (|A_s| + |B_s|)}$.

4 Discussion and Conclusion

The emergence of very high quality registration technology has led to the emergence of segmentation-by-registration strategies where grayscale images are aligned and a known segmentation is warped to the target subject. Since the alignment of images from two individuals inevitably contains some registration error, multi-atlas label fusion techniques have been developed in order to marginalize registration error by employing template subjects with a distribution of anatomy. These methods are fundamentally limited in their ability to identify anatomy that cannot be captured through registration alone. Sabuncu [6] attempts to address this issue by incorporating an image intensity similarity term and pairs labelmap templates with intensity templates. This strategy inevitably requires data from the same modality and is incapable of modeling patient or disease specific intensity changes. Lötjönen [13] recognized this shortcoming and added an intensity classification *after* label fusion, however doing so conflates registration error and anatomical variability. They reported a Dice overlap improvement of 0.01 – 0.02 from the addition of intensity classification.

The method we introduce here is a new type of multi-template classifier fusion that exploits the strengths of both label fusion and statistical classification. The anatomical variation in the population is encoded in a library of input templates. By weighting these appropriately, and using them independently, we're able to employ a learning strategy that identifies areas of mismatch between each template and the target subject. This provides improved classification for diffuse, distributed structures with strong intensity contrast, as well as for structures which are largely isointense, but well defined by their position in the anatomy. Segmentation of the latter is heavily influenced by the template-based prior

probabilities, while the former are accurately identified using the learned subject-specific likelihood functions.

Acknowledgments. This investigation was supported in part by NIH grants R01 RR021885, R01 EB008015, R03 EB008680 and R01 LM010033. The authors thank Dr. Alireza Akhondi-Asl for his implementation of the combined intensity/label fusion algorithm described in [6].

References

1. Vannier, M.W., Butterfield, R.L., Jordan, D., Murphy, W.A., Levitt, R.G., Gado, M.: Multispectral analysis of magnetic resonance images. *Radiology* 154(1), 221–224 (1985)
2. Wells, W.M., Grimson, W.L., Kikinis, R., Jolesz, F.A.: Adaptive segmentation of MRI data. *IEEE Trans. Med. Imaging* 15(4), 429–442 (1996)
3. Van Leemput, K., Maes, F., Vandermeulen, D., Suetens, P.: Automated model-based tissue classification of MR images of the brain. *IEEE Trans. Med. Imaging* 18(10), 897–908 (1999)
4. Rohlfing, T., Russakoff, D.B., Maurer Jr., C.R.: Performance-based classifier combination in atlas-based image segmentation using expectation-maximization parameter estimation. *IEEE Transactions on Medical Imaging* 23(8), 983–994 (2004)
5. Aljabar, P., Heckemann, R.A., Hammers, A., Hajnal, J.V., Rueckert, D.: Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy. *Neuroimage* 46(3), 726–738 (2009)
6. Sabuncu, M.R., Yeo, B.T.T., Van Leemput, K., Fischl, B., Golland, P.: A generative model for image segmentation based on label fusion. *IEEE Trans. Med. Imaging* 29(10), 1714–1729 (2010)
7. Commowick, O.: Design and Use of Anatomical Atlases for Radiotherapy. PhD thesis, University of Nice: Sophia-Antipolis (2007)
8. Warfield, S.K., Zou, K.H., Wells III, W.M.: Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Trans. Med. Imaging* 23(7), 903–921 (2004)
9. Cover, T.M.: Estimation by the nearest neighbor rule. *IEEE Trans. Information Theory* 14(1), 50–55 (1968)
10. Commowick, O., Warfield, S.K.: Incorporating priors on expert performance parameters for segmentation validation and label fusion: a maximum a posteriori staple. *Med. Image Comput. Assist. Interv.* 13(Pt 3), 25–32 (2010)
11. Weisenfeld, N.I., Warfield, S.K.: Automatic segmentation of newborn brain MRI. *Neuroimage* 47(2), 564–572 (2009)
12. Dice, L.R.: Measures of the amount of ecologic association between species. *Ecology* 26, 207–302 (1945)
13. Lötjönen, J.M., Wolz, R., Koikkalainen, J.R., Thurfjell, L., Waldemar, G., Soininen, H., Rueckert, D.: Alzheimer’s Disease Neuroimaging Initiative: Fast and robust multi-atlas segmentation of brain magnetic resonance images. *Neuroimage* 49(3), 2352–2365 (2010)

Anatomical Labeling of the Anterior Circulation of the Circle of Willis Using Maximum a Posteriori Classification

Hrvoje Bogunović, José María Pozo, Rubén Cárdenes, and Alejandro F. Frangi

Center for Computational Imaging & Simulation Technologies in Biomedicine, (CISTIB) – Universitat Pompeu Fabra (UPF) and CIBER-BBN, Barcelona, Spain

Abstract. Automated anatomical labeling of the arteries forming the Circle of Willis is of great interest as facilitates inter-subject comparison required to discover geometric risk factors for the development of vascular pathologies. In this paper, we present a method for anatomical labeling of vessels forming anterior part of the Circle of Willis by detecting the five main vessel bifurcations. The method is first trained on a set of pre-labeled examples, where it learns local bifurcation features as well as global variation in the anatomy of the extracted vascular trees. Then the labeling of the target vascular tree is formulated as maximum a posteriori solution where the classifications of individual bifurcations are regularized by the prior learned knowledge of the tree they span. The method was evaluated by cross-validation on 30 subjects, which showed the vascular trees were correctly anatomically labeled in 90% of cases. The proposed method can naturally handle anatomical variations and is shown to be suitable for labeling arterial segments of Circle of Willis.

1 Introduction

The Circle of Willis is a ring of cerebral vessels that connects the two halves of the anterior circulation with the posterior one and its bifurcations are a common site of aneurysm formation (pathological dilations of vessels) [1]. Analyzing the statistical variation of characteristics of these vessels and bifurcations is of great interest as it can lead to the identification of *geometric risk factors* for the onset of vascular pathologies. Two important operations required for comparing and registering the vasculature between subjects are landmark matching and anatomical labeling. These are tedious and time consuming tasks to be performed manually. Thus, automating them becomes crucial for processing large number of cases.

In this work, we present a method for robust classification of bifurcations in a vascular tree, which we applied to the task of anatomically labeling the anterior part of the Circle of Willis. The anterior part is formed by the three main vessels: internal carotid artery (ICA), middle cerebral artery (MCA) and anterior cerebral artery (ACA). Statistically, 80% of the aneurysms occur along these vessels [1]. In particular, we are interested in anatomical labeling of the following five bifurcations and the vessels connecting them. The bifurcations of ICA

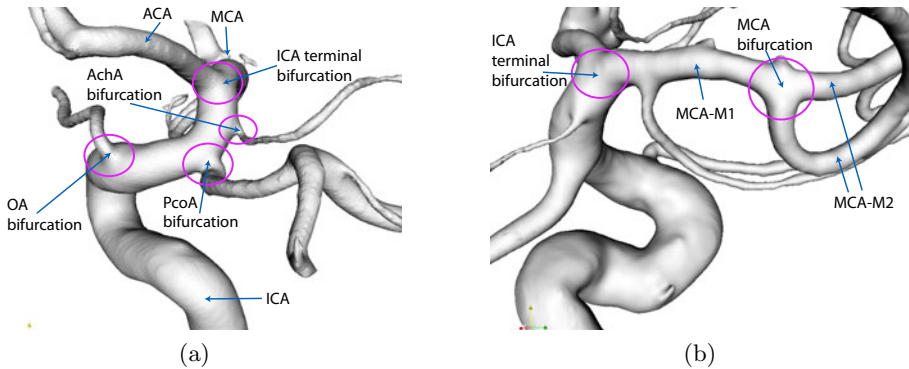


Fig. 1. The arteries of the Circle of Willis and the bifurcations of interest on: (a) internal carotid artery (ICA), (b) middle cerebral artery (MCA)

with ophthalmic artery (OA), posterior communicating artery (PcoA), anterior choroidal artery (AchA) and the terminal ICA bifurcation, plus the principal bifurcation of MCA, which divides the MCA-M1 and MCA-M2 segments (Fig. 1). To that set of bifurcations, throughout the text, we will refer to as *bifurcations of interest* (BoI).

The state of the art for anatomical labeling of vasculature is, in general, sparse. The seminal work was done by Tschirren et al. [2] for anatomical labeling of airway trees that relies on branchpoint matching to a pre-labeled tree that represents population average. Mori et al. in a series of works [3,4] presented a knowledge-based framework for anatomical labeling of bronchial branches based on machine learning and combination optimization. However, their method uses a fixed set of topological constraints and does not seem to be robust to large anatomical variations of the reference tree topology. Airway and bronchial trees are characterized by many similar bifurcations connected by short straight branches. This makes the methods design for them difficult to apply to the task of labeling cerebral vessels where the vessels are in general longer and more curved and the bifurcations more complex. Recently, Mori et al. [5] tuned their approach to a specific task of labeling abdominal arteries but even the authors themselves stated that the application to other organs would be challenging. Bogunović et al. [6] showed that the ICA terminal bifurcation can be successfully classified using support vector machine (SVM) but only a single bifurcation is detected and no tree properties are considered.

The rest of the paper is organized as follows. In section 2, we present the workflow of the method, starting from angiographic 3D image, and formulate the classification of bifurcations in a maximum a posteriori framework. Evaluation of the method on a set of 30 images is presented in section 3. Finally, section 4 discusses the advantages and limitations of the proposed method and concludes the paper.

2 Methods

2.1 Vascular Tree Extraction and Characterization

Segmentations of the cerebral vasculature from 3D images were performed in a fully automated way with a geometric deformable model called Geodesic Active Regions (GAR) [7]. From the segmented vasculature, fast topological thinning based on collapsing fronts using [8] was applied to obtain a rough estimate of the underlying topology of the vessel tree. Due to image acquisition noise and inaccuracies in vessel segmentation, touching vessels effect can appear, thus causing the extracted topology to form a connected graph containing cycles and not a tree. However, the graph's end-points do correspond to the root and the terminal leaves of the vascular tree. Root was taken to be the end-point with the maximal associated radius at the lowest axial plane which, in our images, corresponded to the ICA entering the field of view. Then, the set of accurate centerlines is obtained by backtracking along the minimal cost path from the end-points toward the root using [9], which is available within the open-source library VMTK [10]. These centerlines now form a rooted tree with the edges directed away from the root in accordance with the blood flow (Fig. 2(a)).

Bifurcation Feature Vector. Each bifurcation of the tree is characterized using the method of Antiga [11], available in VMTK, which relies on objective criteria for defining the origin of a bifurcation and bifurcation vectors of the parent branch and the two daughter branches. The two daughter branches are differentiated by their radius: Larger daughter branch and smaller daughter branch. We chose the following collection of features to quantify the geometry of a bifurcation (Fig. 2(b)), which form a 21 dimensional, scale invariant, bifurcation feature vector:

- Sagittal, axial and coronal-components of the normal vector of the bifurcation plane (3).
- Sagittal, axial and coronal-components of the three bifurcation vectors (9).

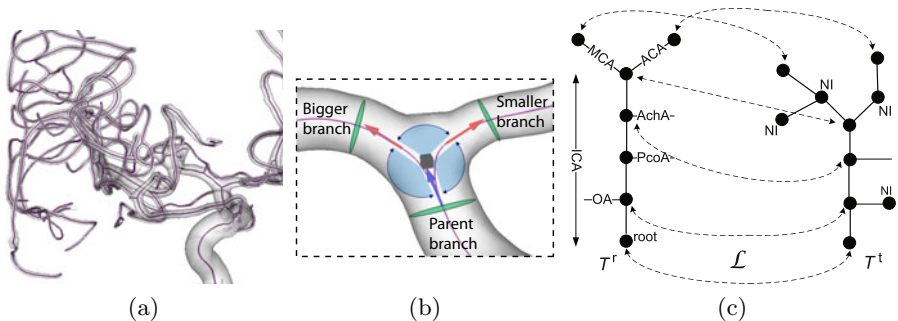


Fig. 2. Elements of the method's workflow: (a) Centerline extraction. (b) Bifurcation features. (c) Example of a labeling \mathcal{L} (dotted arrows) of a target tree T^t (right) based on a reference tree T^r with the bifurcations of interest (left).

- Angles between each pair of bifurcation vectors (3).
- Ratios of mean vessel radii between each pair of bifurcation branches (3).
- Ratios of mean vessel radii between each bifurcation branch and the root branch (ICA) of the tree (3).

Tree Features. To characterize the extracted tree, to each pair of bifurcations we associate an inheritance relationship: *ancestor-offspring* or *sibling*. In addition, a probability is assigned that a particular bifurcation is present in the tree. In general, one could easily also include features of the vessels connecting the pair of bifurcations, like their mean length, curvature, torsion or tortuosity.

2.2 Bifurcation Classification

Extracted vascular tree can be considered as Attributed Relational Tree (ART) which has attributes in a form of feature vectors associated to its vertexes and edges. ART is defined as

Definition 1. *Rooted Attributed Relational Tree is a quadruple $T = (V, E, A, r)$, where V is the vertex set, E is the edge set, r is root and A is the attribute set that contains unary attribute a_i attaching to each vertex $v_i \in V$, and a binary attribute $a_{i,j}$ attaching to each edge $e_k = (n_i, n_j) \in E$.*

The classification of bifurcations is based on the availability of a reference ART T^r , created from a representative sample of a population, having BoI as its vertices $V^r = \{v_i^r : 1 \leq i \leq M\}$ together with an entire set of sample bifurcation feature vectors as $\{a_i^r\}$, while edge features $\{a_{i,j}^r\}$ are currently not used. Such a reference ART is normally created from a training set of trees. The target ART T^t corresponds to the extracted vascular tree, having its bifurcations as vertices $V^t = \{v_j^t : 1 \leq j \leq N\}$ and the bifurcation feature vectors as $\{a_j^t\}$. Then, on the target tree we define a labeling process $\mathcal{L} : V^t \rightarrow V^r \cup \{\text{NI}\}$, where the label NI represents a bifurcation which is *not of interest* and is not present in the reference ART (Fig. 2(c)).

We are interested in estimating the probability $P(\mathcal{L}|T^t, T^r)$ of \mathcal{L} being correct, and finding the mode of this posterior distribution. Thus, the problem can be formulated as finding a labeling \mathcal{L}^* that satisfies the maximum a posteriori (MAP) solution

$$P(\mathcal{L}^*|T^t, T^r) > P(\mathcal{L}|T^t, T^r) \quad \forall \mathcal{L} \neq \mathcal{L}^*, \tag{1}$$

$$P(\mathcal{L}|T^t, T^r) \propto P(T^t|\mathcal{L}, T^r)P(\mathcal{L}|T^r), \tag{2}$$

where the constant denominator term in Eq. 2 has been left out for brevity.

The *prior* term $P(\mathcal{L}|T^r)$ presents our knowledge-based expectations about the topology and tree properties of the labeled target tree and regularizes the classification based on bifurcation features. The *likelihood* term, assuming that the feature vectors of each bifurcation $\{a_i\}$ are statistically independent from each other, can be written as

$$P(T^t|\mathcal{L}, T^r) = \prod_{i=1}^N P(a_i^t|\mathcal{L}(v_i^t), T^r). \tag{3}$$

Likelihood Estimation. To estimate $P(a_i^t | \mathcal{L}(v_i^t))$ i.e. a likelihood that a bifurcation with known label $\mathcal{L}(v_i^t)$ has the feature vector a_i^t , we employed the SVM with probability estimates available in LIBSVM open-source library [12]. A non-linear model based on radial basis function (RBF) kernel was used. For the multi-class problem, several binary SVM were created in a one-against-one strategy where binary SVMs, for all pairs of classes, were trained on the data and the results for each pair accumulate to yield the final probability distribution.

MAP Estimation. Unfortunately MAP estimation is extremely computationally expensive. Testing for all possible labelings \mathcal{L} is not feasible. However, the *prior* term $P(\mathcal{L}|T^r)$ is positive only for a small subset of all combinations i.e. only for those for which the topology corresponds to the one of the reference tree T^r . This happens when \mathcal{L} is an isomorphism from a subgraph of T^t to a subdivision of T^r .

A standard algorithm to find all isomorphic mappings between two graphs (trees) is based on building their association graph and then finding all *maximal cliques* of such undirected graph. Association graph $G = (V^a, E^a)$ is built from T^r and T^t , where nodes are denoted with a pair of indices $V^a = \{v^a \equiv (v_i^r, v_j^t)\} \equiv V^r \times V^t$, with the following rule. Edge $(v_{i,j}^a, v_{k,l}^a)$ is created only if v_i^r and v_k^r have the same relationship (*ancestor-offspring* or *sibling*) as do v_j^t and v_l^t , for $i \neq k$ and $j \neq l$. We have now extended the above rules by adding the label NI to V^t , hence $V^a = V^r \times (V^t \cup \{\text{NI}\})$ with the special rule that NI is in all relationships with any other vertex of V^a including itself.

Finding maximal cliques (in our case they will always be of size N) in an undirected graph is in general NP-complete problem and we used Bron-Kerbosch algorithm [13], that tries to reduce the size of search space, to find the maximal cliques.

3 Results

We evaluated the proposed methodology on a dataset consisting of images from 30 subjects, acquired with 3D rotational angiography (3DRA), where contrast injection was carried out to enhance the vessels comprising the anterior cerebral circulation of either left or right hemisphere. Images were reconstructed with a 256^3 matrix having a voxel size of $0.29 \times 0.29 \times 0.29$ (mm).

All images were successfully segmented and had their centerlines and tree topology extracted. The obtained trees were containing bifurcations and trifurcations. Since two close bifurcations are sometimes reconstructed as a trifurcation, we decided to always treat trifurcations as two bifurcations sharing the same parent branch and the larger daughter branch.

The success of bifurcation classification and hence the anatomical labeling was evaluated using cross-validation. One observer manually labeled the five BoI on each extracted tree. All the other bifurcations were assigned the label NI. Leave-one-out cross-validation was then run, where iteratively $K = 30$ times, one case is used for testing while $K - 1$ cases were selected to form the training set from which the bifurcation vectors, reference tree topology and its probability of appearance were learned.

To evaluate the contribution of the regularizing *prior* term, we performed the above validation for three method variants: the first variant is a natural extension of [6] and is relying just on the SVM classification based on the bifurcation feature vectors with no regularization. The second variant has the regularization by topology added (SVM+TP), and can be considered as an adaptation of [5]. Finally, the third variant is the proposed one, which includes the prior term (SVM+TP+AP), where the probability of appearance of a particular anatomical variability coming from the missing BoI is learned from the training set. The cross-validation results are presented in Table 1.

Table 1. Performance of the three method variants. Scores for successfully labeling the complete vessel tree and detecting the individual bifurcations are given.

Method	Vessel tree labeling rate	Bifurcation detection rate						
		OA	PcoA	AchA	ICA	MCA	NI	
SVM	~[6]	0.7	1.00	1.00	1.00	0.96	0.89	0.85
SVM+TP	~[5]	0.8	1.00	1.00	1.00	0.96	0.82	0.94
SVM+TP+AP		0.9	1.00	1.00	1.00	1.00	0.96	0.91

As seen from the first row, PcoA, AchA and OA bifurcations were well classified just based on their feature vectors. Gradually adding the regularization terms improved the tree labeling accuracy each time by 10%. The proposed method (SVM+TP+AP) outperformed similar state of the art methods and had success rate of 90% for correctly labeling the complete vessel tree. The error occurred in three cases, when MCA bifurcation was detected as a NI and vice versa. A sample of correct anatomical labeling results is shown in Fig. 3 where the surface area of each corresponding vessel segment is labeled.

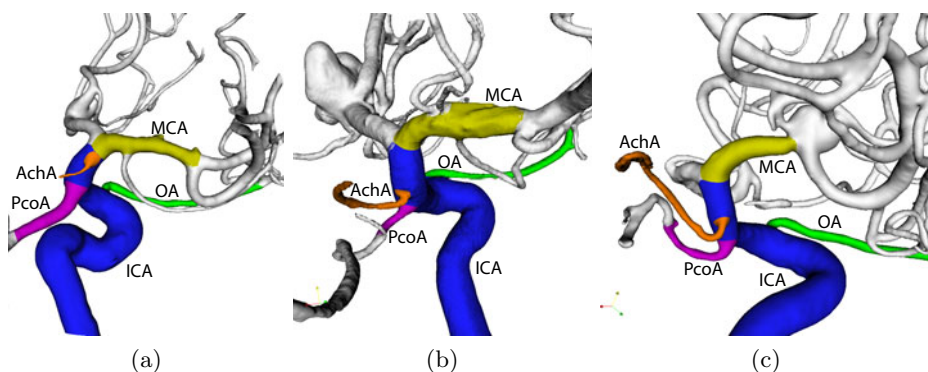


Fig. 3. Anatomically labeled vascular trees. Denoted vessels: ICA (blue), OA (green), PcoA (violet), AchA (orange), MCA (yellow)

As the method is formulated in a probabilistic framework the candidate labelings can be ranked by their estimated probabilities. The method always selects the most probable one (the highest ranked one). However, in the case that selected labeling is visually observed to be incorrect, this enables us to select the next most probable candidate in the ranked list. For the three cases where the method assigned the incorrect labels, we searched for the correct labeling in the ranked list (Table 2). We can observe that the correct labeling appears as the second or at most the third most probable candidate.

Table 2. The three cases for which the labeling provided by the proposed method (SVM+TP+AP) was incorrect. Position in the list of candidate labelings ranked by their probabilities is given for the selected and the correct labeling.

	position (probability)	
	selected	correct
error case 1	1 (0.61)	2 (0.20)
error case 2	1 (0.39)	2 (0.26)
error case 3	1 (0.36)	3 (0.15)

4 Conclusion

We presented a method for anatomical labeling of the main vessels forming anterior part of the Circle of Willis. From a set of prelabeled examples, the method is able to learn local bifurcation features as well as global variation in tree topology and their probabilities of occurrence. Focusing just on BoI makes it insensitive to the number of bifurcations that are extracted in between them. This is important as in general, total number of bifurcations and vessels largely varies between images as it heavily depends on the contrast and image resolution during acquisition as well as accuracy of segmentation method applied.

The current main limitation is that topologically correct candidates are obtained by computing maximal cliques of association graph, which is a NP-complete problem. For our application, this was computationally feasible as in general trees could be pruned in a preprocessing step down to $N \sim 20$ prospective candidates. In the future, we plan to search for more efficient way of finding isomorphic mappings and also extend the anatomical labeling to the entire Circle of Willis.

Acknowledgments. This work has been partially funded by the CENIT programme of CDTI (cvREMOD CEN-20091044) and VPH-NoE project from FP7 of the European Union (nr 223920). The work of H.B. is supported by the FI programme of AGAUR. The work of R.C. is supported by the Beatriu de Pinós, programme of AGAUR. A.F.F. is partially funded by the ICREA-Academia programme.

References

1. Brisman, J.L., Song, J.K., Newell, D.W.: Cerebral aneurysms. *N. Engl. J. Med.* 355(9), 928–939 (2006)
2. Tschirren, J., McLennan, G., Palagyi, K., Hoffman, E.A., Sonka, M.: Matching and anatomical labeling of human airway tree. *IEEE Trans. Med. Imag.* 24(12), 1540–1547 (2005)
3. Mori, K., Hasegawa, J., Suenaga, Y., Toriwaki, J.: Automated anatomical labeling of the bronchial branch and its application to the virtual bronchoscopy system. *IEEE Trans. Med. Imag.* 19(2), 103–114 (2000)
4. Mori, K., Ota, S., Deguchi, D., Kitasaka, T., Suenaga, Y., Iwano, S., Hasegawa, Y., Takabatake, H., Mori, M., Natori, H.: Automated anatomical labeling of bronchial branches extracted from CT datasets based on machine learning and combination optimization and its application to bronchoscope guidance. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) *MICCAI 2009*. LNCS, vol. 5762, pp. 707–714. Springer, Heidelberg (2009)
5. Mori, K., Oda, M., Egusa, T., Jiang, Z., Kitasaka, T., Fujiwara, M., Misawa, K.: Automated nomenclature of upper abdominal arteries for displaying anatomical names on virtual laparoscopic images. In: Liao, H., Edwards, E., Pan, X., Fan, Y., Yang, G.Z. (eds.) *MIAR 2010*. LNCS, vol. 6326, pp. 353–362. Springer, Heidelberg (2010)
6. Bogunovic, H., Pozo, J.M., Cardenas, R., Frangi, A.F.: Automatic identification of internal carotid artery from 3DRA images. In: *Proc. IEEE EMBC*, pp. 5343–5346. IEEE Press, Los Alamitos (2010)
7. Bogunovic, H., Pozo, J.M., Villa-Uriol, M.C., Majoie, C.B.L.M., van den Berg, R., Gratama van Andel, H.A.F., Macho, J.M., Blasco, J., San Roman, L., 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, 210–222 (2011)
8. Cardenas, R., Bogunovic, H., Frangi, A.F.: Fast 3D centerline computation for tubular structures by front collapsing and fast marching. In: *Proc. IEEE ICIP*, pp. 4109–4112. IEEE Press, Los Alamitos (2010)
9. Antiga, L., Ene-Iordache, B., Remuzzi, A.: Computational geometry for patient-specific reconstruction and meshing of blood vessels from MR and CT angiography. *IEEE Trans. Med. Imag.* 22(5), 674–684 (2003)
10. Antiga, L., Steinman, D.A.: VMTK: the Vascular Modeling Toolkit (2011), <http://www.vmtk.org>
11. Antiga, L., Steinman, D.A.: Robust and objective decomposition and mapping of bifurcating vessels. *IEEE Trans. Med. Imag.* 23(6), 704–713 (2004)
12. Chang, C.C., Lin, C.J.: LIBSVM: a library for support vector machines (2011), <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
13. Bron, C., Kerbosch, J.: Algorithm 457: finding all cliques of an undirected graph. *Comm. ACM* 16(9), 575–577 (1973)

Automatic Multi-organ Segmentation Using Learning-Based Segmentation and Level Set Optimization

Timo Kohlberger¹, Michal Sofka¹, Jingdan Zhang¹, Neil Birkbeck¹,
Jens Wetzl¹, Jens Kaftan², Jérôme Declerck², and S. Kevin Zhou¹

¹ Image Analytics and Informatics, Siemens Corporate Research, Princeton, NJ, USA

² Molecular Imaging, Siemens Healthcare, Oxford, UK

Abstract. We present a novel generic segmentation system for the fully automatic multi-organ segmentation from CT medical images. Thereby we combine the advantages of learning-based approaches on point cloud-based shape representation, such a speed, robustness, point correspondences, with those of PDE-optimization-based level set approaches, such as high accuracy and the straightforward prevention of segment overlaps. In a benchmark on 10–100 annotated datasets for the liver, the lungs, and the kidneys we show that the proposed system yields segmentation accuracies of 1.17–2.89mm average surface errors. Thereby the level set segmentation (which is initialized by the learning-based segmentations) contributes with an 20%-40% increase in accuracy.

1 Introduction

Discriminative segmentation approaches have proven to give reliable, fully-automatic, and fast detections of anatomical landmarks within volumetric images, as well as the accurate determination of organ boundaries, such as of the inner and outer walls of the heart, [10], or of the liver, [4]. Usually, the segmenting surface is represented by relatively low number of explicit control points, such as they are used in Active Shape Models.

Besides restrictions in topology, other well-known disadvantages of point cloud-based shape representations are the dependence of the local detailedness on the local density of control points. The latter often are non-homogeneously distributed across the shape boundary, and thus yield varying levels of segmentation accuracy. Level set-based shape representations, see [1] and the references therein, on the other hand, provide a well-known mean to encode segment boundaries at a homogeneous resolution, with simple up- and down-sample schemes. Moreover, in the case of multiple objects, the detection and formulation of constraints to prevent overlaps between adjacent segment boundaries can be achieved much simpler by a level set representation where signed distance functions are employed.

In the following we will present a fully automatic segmentation system for multiple organs on CT data, that combines the advantages of both segmentation

approaches and their employed shape representations in an optimal manner. In particular the point-to-point correspondences, which are estimated during the learning-based segmentation, will be preserved in the level set segmentation. With regards to the latter, we will present novel terms which allow to impose region-specific geometric constraints between adjacent boundaries.

In Section 2 we first give an outline of the learning-based detection and segmentation stages, all of which have been published in the mentioned citations. After a brief description on the conversion of their output meshes to implicit level set maps, we start to build an energy-based minimization approach for multiple organs as a segmentation refinement stage. In the experimental Section 3, we first show the impact of the new constraints at qualitative examples, and finally evaluate the overall improvement of the level set refinement stage over the detection-based results at 10–100 annotated cases for the liver, lungs and kidneys.

2 Approach

2.1 Anatomical Landmark Detection and Learning-Based Segmentation of Organ Boundaries

We initialize the multi-region level set segmentation from explicitly represented boundary surfaces stemming from an existing learning-based detection framework, which itself consists of several stages. In the first stage, a landmark detection system estimates key organ landmarks ranging from the abdominal to upper body region, see 5 for more details. These landmarks then serve as initializations for a hierarchical bounding box detection system based on Marginal Space Learning 10 in connection with Probabilistic Boosting Trees 9. The latter yields bounding box estimates for the liver, the left and right lung, the heart, and the kidneys. In the third stage, organ-specific boundary detectors are employed to evolve the correct organ boundaries on a coarse scale and in subsequently on a fine scale, see 4. In addition, PCA-based statistical shape models are used to regularize the boundary shape on the coarse resolution. Thereby, segment boundaries of each organ are represented by a triangulated mesh, i.e. a connected point-cloud as being used in Active Shape Models.

2.2 From Meshes to Zero-Crossings of Signed Distance Maps

Although the learning-based segmentation sub-system already provides good individual organ segmentations, see Fig 2(a) for example, they usually exhibit small overlaps between adjacent organ boundaries, or gaps where the true organ boundaries coincide. Given representations of only the adjacent segments' boundaries, those deficiencies are difficult to detect and remove. Instead we initialize signed distance functions $\phi_i : \mathbb{R}^3 \Rightarrow \mathbb{R}$ from each of the result meshes C_i , for i, \dots, N organs by employing a fast mesh voxelization algorithm. The boundary information then is encoded implicitly in the zero-crossings of the ϕ_i , i.e.

$C_i := \{\mathbf{x} \mid \phi_i(\mathbf{x}) = 0, |\nabla\phi| = 1\}$, with $|\nabla\phi| = 1$ denoting the so-called distance-property, and $\phi > 0$ inside the object and < 0 outside, see [1] and the references therein. The distance functions are discretized on a regular grid, which, in the following, is assumed to be the same for all organs. Furthermore, we employ a narrow-banded level set scheme, which maintains the distance-property in a small narrow-band of ± 2 voxels from the zero crossing. In addition to the distance functions, we still keep the mesh points, denoted by $\{\mathbf{p}_i\}$, and are tracking them along with the evolving zero crossing as described in [3], since they provide point-wise correspondences to the mean shape of the PCA model employed in the preceding learned-based boundary detection step.

2.3 Data and Smoothness Term of the Level Set Approach

Having the triangulated boundary meshes $\{C_i\}$ of the detection stages transferred to distance functions $\{\phi_i\}$, detecting and removing local overlaps and gaps between them can be realized much easier. The ultimate target of course is to find the correct separating boundary between two neighboring organs. To that end, in the following, we propose a level set segmentation approach which not only refines the segmentation boundary accuracy, removes local overlaps and gaps, but also finds the true separating boundary given that enough image information is available.

For each organ O_i , this refining level set segmentation is realized by employing gradient descent iteration to converge to a minimum of an associated energy functional $E_i(\phi_i)$, given the initial distance maps as starting points, see [1] and the references therein. As data-dependent energy term, we here employ

$$E_p(\phi) = -\alpha \int_{\Omega} H(\phi) \log p_{in}(I(\mathbf{x})|\phi) + (1 - H(\phi)) \log p_{out}(I(\mathbf{x})|\phi) d\mathbf{x},$$

with H denoting the Heaviside step function, and $p_{in/out}$ referring to non-parametric probability estimates of the intensities inside and outside, respectively, of the current segment ϕ using a Parzen density estimator with a Gaussian kernel, see [2] for further details, and α being a constant weight. In order to add robustness against noisy data, we furthermore incorporate the boundary smoothness regularization term

$$E_c(\phi) = \int_{\Omega} \gamma_{l(\mathbf{x})} |\nabla H(\phi)| d\mathbf{x}, \quad \text{with } l(\mathbf{x}) = \arg \min_{i=1,\dots,N} \|\mathbf{x} - \mathbf{p}_i\|_{\mathcal{L}_2}, \quad (1)$$

which employs a weight $\gamma_{l(\mathbf{x})}$ that varies with the location on the boundary. The latter is realized by assigning fixed weights $\{\gamma_i\}$ to each of the correspondence points $\{\mathbf{p}_i\}$ on the mean shape of the PCA shape model, which then are tracked along during the zero-crossing evolution, see [3] for more details.

2.4 Disjoint Constraint to Remove Overlaps

Let C_A^0 and C_B^0 be detected boundaries of the learning-based stage in the case of two adjacent organs A and B , and assume that C_A^0 and C_B^0 imperfectly overlap

each other to a certain degree, see, e.g., Fig. [1\(a\)](#). By representing these two surfaces using signed distance functions ϕ_A^0 and ϕ_B^0 , locations $\tilde{\mathbf{x}}$ inside the overlapping region are exclusively characterized by $\phi_A(\tilde{\mathbf{x}}) > 0$ and $\phi_B(\tilde{\mathbf{x}}) > 0$, and thus provide a much simpler overlap indicator compared to any other based on an explicit shape representation. Subsequently, additional energy terms which explicitly penalize overlaps usually are of the form

$$E_o(\phi_A, \phi_B) := \int_{\Omega} H(\phi_A(\mathbf{x})) H(\phi_B(\tilde{\mathbf{x}})) \phi_B(\tilde{\mathbf{x}}) d\mathbf{x} \quad (2)$$

where the first product in the integrand is unequal zero only inside the overlap regions, see Fig. [1\(b\)](#). In addition to similar terms such as proposed in [\[6\]](#), we propose to also multiply with the second distance function ϕ_B which makes E_O smoother at the presence of small overlaps, and thereby decreases oscillations during gradient descent. The corresponding energy gradient reads (cf. [\[7\]](#)):

$$\partial\phi_A/\partial t = -\partial E_o/\partial\phi_A = -\delta_\epsilon(\phi_A) H_\epsilon(\phi_B(\tilde{\mathbf{x}}))\phi_B(\tilde{\mathbf{x}}). \quad (3)$$

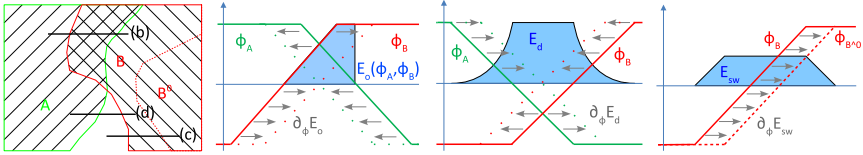


Fig. 1. Imposing geometric constraints to remove overlaps (b) and gaps (c) from an existing segmentation, while controlling the deviation from a given shape locally (d). (Note that (b)–(d) are 1D-cuts along the black lines in (a)).

2.5 Local Proximity Constraint to Fill Gaps

With regards to removing erroneous gaps between adjacent segmentation boundaries, we add the following energy to the total energy functional:

$$E_d(\phi_A, \phi_B) := \frac{1}{2} \int_{\Omega} \beta_l(\mathbf{x}) (\phi_A(\mathbf{x}) + \phi_B(\tilde{\mathbf{x}}) + D)^2 d\mathbf{x} \quad (4)$$

with $D = 0$ for the time being, and $\{\beta_i\}$ being correspondence points-bound weights with $\beta_i = 0$ at points where no boundary coincidence ought to be enforced, and $\beta_i > 0$ at locations where boundaries of A and B ought to coincide. As illustrated in Fig. [1\(c\)](#), ϕ_A and ϕ_B cancel each other out if their zero crossings coincide and thus the integrand becomes zero. As an extension, one can enforce the two boundaries to not touch but stay in a predefined distance $D > 0$ from each other. The gradient descent PDE of E_d w.r.t. ϕ_A reads:

$$\partial\phi/\partial t = -\partial E_d/\partial\phi = -\beta_l(\mathbf{x}) (\phi_A(\mathbf{x}) + \phi_B(\tilde{\mathbf{x}}) + D) \quad (5)$$

which shows that of ϕ_A increases at locations where $\phi_B < D$, and thus expands its representing boundary, and decreases at locations where $\phi_B > D$, i.e. shrinks the boundary.

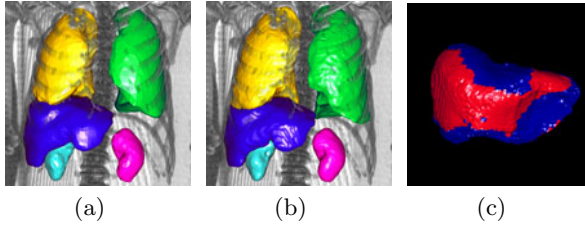


Fig. 2. (a) Segmentation result of the machine learning system. (b) Refined segmentations after applying level set segmentation to (a). (c) Visualization of the manually set local weights $\{\omega_i^{out}\}$ of the outward template constraint (red: 50, blue: 0.2). These weights are bound to point-based shape correspondences in relation to a fixed model shape, and thereby allow for a region-specific control over the different geometrical constraints during the level set segmentation.

2.6 Template Constraint to Control Deviation of the Refinement

Finally, we added a third geometric term, which ensures that the level set result is sufficiently similar to learning-based contour, that is, the refined boundary is sought only in the vicinity of its initialization. To that end, we use the term

$$E_{sw}(\phi, \phi_P) := \frac{1}{2} \int_{\Omega} \omega_l^{in}(\mathbf{x}) H(\phi_P(\mathbf{x}) - \phi(\mathbf{x})) + \omega_l^{out}(\mathbf{x}) H(\phi(\mathbf{x}) - \phi_P(\mathbf{x})) \, d\mathbf{x} . \quad (6)$$

which is an extension of the approach shown in [8] in the sense that it applies region-specific weights $\{\omega_i^{in}\}$ to the shape dissimilarity measure between the current ϕ and the prior shape ϕ_P (which is the initial one here), as well as applying different weights for deviations outside or inside of C_P (we refer to C_P as “template” shape in the following). See Fig. 2(c) for a local weight map. Technically, note that the first term of the integrand is non-zero only if the zero-crossing of ϕ resides inside the zero-crossing of ϕ_P , that is the current boundary C is smaller than the prior boundary C_P , see Fig. 1(d). Vice-versa, the second term measures local expansions relative to C_P , by becoming non-zero only where $\phi(\mathbf{x}) > \phi_P(\mathbf{x})$.

The corresponding energy gradient clearly shows that the proposed energy term has the desired effect:

$$\partial\phi/\partial t = -\partial E_{sw}/\partial\phi = \omega_l^{in}(\mathbf{x}) \delta_{\epsilon}(\phi_P(\mathbf{x}) - \phi(\mathbf{x})) - \omega_l^{out}(\mathbf{x}) \delta_{\epsilon}(\phi(\mathbf{x}) - \phi_P(\mathbf{x})) , \quad (7)$$

i.e. increasing ϕ at locations where $\phi < \phi_P$, and decreasing it in the opposite case.

2.7 Interleaved Multi-energy Minimization

Finally, all of the proposed energy terms are combined into energy minimizations for each organ $O_{i=1,\dots,N}$:

$$\min_{\phi_i} E_p(\phi_i) + E_c(\phi_i) + \sum_{j \in \mathcal{N}_i(j)} E_o(\phi, \phi_j) + \sum_{j \in \mathcal{P}_i(j)} E_d(\phi, \phi_j) + E_{sw}(\phi_i, \phi_i^0) , \quad (8)$$

which are mutually coupled by the disjoint and proximity terms (\mathcal{N}_i : indices of organs adjacent to O_i , \mathcal{P}_i : indices of organs with which O_i shares a mutual proximity constraint). Consequently, minimizers $\{\tilde{\phi}_i\}$ of these individual energies depend on each other. To that end, we found interleaved gradient descent iterations to yield the desired segmentation improvements in practice. Specifically, we carry out a descent along the negative gradients of the N per-organ energies in lockstep, while using the segmentation results $\{\phi_i^{t-1}\}$ of the previous joint iteration to compute the coupled energy gradients $\partial E_i(\phi_i; \{\phi_i^{t-1}\})/\partial \phi_i$. The descent for a particular energy is terminated if a given maximum number of iterations has been reached, or if the maximum norm of its gradient falls below a given threshold, i.e. the segmentation boundary ϕ_i changes less than a chosen tolerance.

3 Experimental Evaluation

3.1 Parameter Selection

In a first series of experiments, we studied the effect of the proposed new energy terms qualitatively on a few data sets. Thereby we also manually selected all the involved weights in order to achieve optimal results on a small set of test cases. Specifically, we weighted the data term E_p with a factor of 2 for all organs and set the lowest weight γ_i in the smoothness term E_c to 0.7 at locations where a high curvature is desired (such as at the lung tips), and to a value of 1.5 where low curvatures are to be enforced. For the disjoint energy term E_o we found a weighting of 1000 to remove any existing overlaps while not producing any oscillations. The proximity term E_d turned out to yield the desired results when setting β_i to 10 at correspondence points where it ought to be active and to zero elsewhere. For the template energy E_{sw} the inward and outward deviation-penalizing weights ω^{in} and ω^{out} were set according to the overall quality and robustness of the learning-based boundary result. Specifically, at locations where the latter tends to under-segment, such as at the lower tips of the lung wings, ω^{out} was set to the low value of 0.2 in order to allow the level set result to deviate outwards. Vice-versa, at locations where the learning-based stage tends to over-segment, such as at the lower side of the liver ω^{in} is given a low value. Finally, at locations where the learning-based stage already yields highly accurate results, such as adjacent to the ribs for the liver and lung boundaries, both ω^{in} and ω^{out} were set to high values of 50 in order to bind the level set-based boundary close to it. Fig. 3(c) visualizes the weights $\{\omega_i^{out}\}$ for a final segmentation of the liver.

3.2 Accuracy Benchmark

In a next step, we benchmarked the overall system at manually annotated data sets in order to study the overall accuracy improvement yielded by level set refinement. Test sets for the different organs listed in Table 1 were drawn randomly from a set of 434 annotated CT cases consisting of low to high contrast scans, and an average voxel spacing of 1.02/1.02/2.16mm (x/y/z). Cases not drawn for

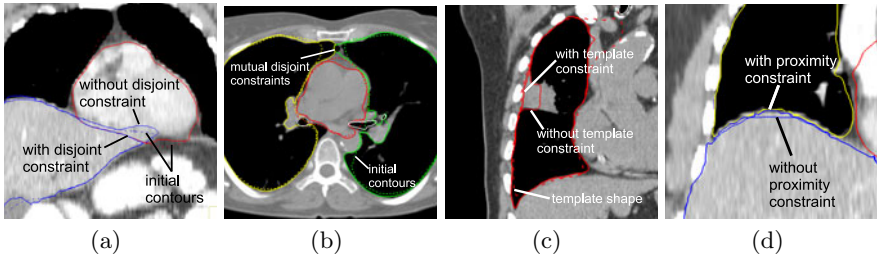


Fig. 3. Effect of the different proposed geometric constraints. Whereas the disjoint constraint (a)+(b) can be used to remove overlaps between initial segmentations, the template constraint (c) can bind the level set zero-crossing to the initial one in a location-specific manner. With the proximity constraint (d), coincidence of shared boundaries can be imposed locally.

testing were used to train the learning-based pipeline, which were 313 boundary annotations for the liver and about 130 annotations for the other organs.

After the the joint gradient descent had converged w.r.t. each energy of the level set segmentation, the final meshes were extracted from the level set maps via Marching Cubes. As error measure we first computed the shortest Euclidean distances between each result mesh and its corresponding annotated mesh at every vertex of the former as well as every vertex of the latter, and then averaged all such distances. The results in Table 1 show that the presented segmentation system yields state-of-the-art accuracies ranging from 1.17mm average surface error for left kidney to 2.89mm for the liver. Thereby the proposed level set segmentation contributes with an improvement of 20% for the liver, 30% for the left lung and left kidney, and 40% for the right lung and right kidney. Average run-times for the full detection pipeline is 2–3 min, which includes about 1 min for the only coarsely multi-threaded level set stage on a 2.0 GHz eight-core Xeon machine.

Table 1. Symmetric surface errors using machine learning-based segmentation and after applying level set-based refinement

Sym. surface error (mm)	Mean	Std.dev.	Median	Worst	80% # cases
Liver, learning-based only	3.5	1.7	3	4.0	100
Liver, level set-refined	2.9	1.7	2.6	3.6	100
Left lung, learning-based only	2.1	0.5	1.9	2.5	60
Left lung, level set-refined	1.5	0.3	1.4	1.7	60
Right lung, learning-based only	2.7	0.9	2.4	3.0	60
Right lung, level set-refined	1.6	0.6	1.5	1.8	60
Right kidney, learning-based only	1.9	0.9	1.8	2.0	10
Right kidney, level set-refined	1.1	0.9	0.8	3.9	10
Left kidney, learning-based only	1.9	1.0	1.7	2.3	10
Left kidney, level set-refined	1.3	1.0	1.0	1.9	10

4 Conclusion

In our experiments we found the proposed algorithm to combine the robustness and speed of a machine learning approach with the high accuracy and advantageous distance map representation of a level set approach. Furthermore, the novel level set constraints allow to impose region-specific geometrical priors in the refinement stage. Yet the involved localized weights, as well as the global ones, need to be set manually in one or more parameter tuning sessions. One approach to automatize this step could be to minimize a sum energies in Equ. (8) over a set of fixed shapes gained from the learning-based stage with respect to these parameters. Finally, our experimental results show state-of-the-art accuracy and robustness of the proposed algorithm for five different organs on various unseen data sets.

References

1. Chan, T.F., Vese, L.A.: Active Contours without Edges. *IEEE Trans. on Image Processing* 10(2), 266–277 (2001)
2. Cremers, D., Rousson, M., Deriche, R.: A Review of Statistical Approaches to Level Set Segmentation: Integrating Color, Texture, Motion and Shape. *IJCV* 72(2), 195–215 (2007)
3. Kohlberger, T., Uzunbas, G., Alvino, C., Kadir, T., Slosman, D., Funke-Lea, G.: Organ Segmentation with Level Sets Using Local Shape and Appearance Priors. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) *MICCAI 2009*. LNCS, vol. 5762, pp. 34–42. Springer, Heidelberg (2009)
4. Ling, H., Zhou, S.K., Zheng, Y., Georgescu, B., Suehling, M., Comaniciu, D.: Hierarchical, Learning-based Automatic Liver Segmentation. In: *CVPR 2008*, pp. 1–8. IEEE Press, New York (2008)
5. Liu, D., Zhou, S.K., Bernhardt, D., Comaniciu, D.: Search Strategies for Multiple Landmark Detection by Submodular Maximization. In: *CVPR 2010*, pp. 2831–2838. IEEE Press, New York (2010)
6. Mansouri, A.-R., Mitiche, A., Vázquez, C.: Multiregion Competition: A Level Set Extension of Region Competition to Multiple Region Image Partitioning. *Computer Vision and Image Understanding* 101(3), 137–150 (2005)
7. Paragios, N.: A Variational Approach for the Segmentation of the Left Ventricle in Cardiac Image Analysis. *IJCV* 50(3), 345–362 (2002)
8. 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)
9. Tu, Z.: Probabilistic Boosting-Tree: Learning Discriminative Models for Classification, Recognition, and Clustering. In: *ICCV 2005*, vol. 2, pp. 1589–1596. IEEE Press, New York (2005)
10. Zheng, Y., Barbu, A., Georgescu, B., Scheuering, M., Comaniciu, D.: Four-Chamber Heart Modeling and Automatic Segmentation for 3-D Cardiac CT Volumes Using Marginal Space Learning and Steerable Features. *IEEE Trans. on Medical Imaging* 27(11), 1668–1681 (2008)

Pose-Invariant 3D Proximal Femur Estimation through Bi-planar Image Segmentation with Hierarchical Higher-Order Graph-Based Priors

Chaohui Wang^{1,2}, Haithem Boussaid^{1,2}, Loic Simon¹,
Jean-Yves Lazenec³, and Nikos Paragios^{1,2,*}

¹ Laboratoire MAS, Ecole Centrale de Paris, France

² Equipe GALEN, INRIA Saclay - Île de France, Orsay, France

³ Centre Hospitalier Universitaire Pitié Salpêtrière, Paris, France

Abstract. Low-dose CT-like imaging systems offer numerous perspectives in terms of clinical application, in particular for osteoarticular diseases. In this paper, we address the challenging problem of 3D femur modeling and estimation from bi-planar views. Our contributions are threefold. First, we propose a non-uniform hierarchical decomposition of the shape prior of increasing clinical-relevant precision which is achieved through curvature driven unsupervised clustering acting on the geodesic distances between vertices. Second, we introduce a graphical-model representation of the femur which can be learned from a small number of training examples and involves third-order and fourth-order priors, while being similarity and mirror-symmetry invariant and providing means of measuring regional and boundary supports in the bi-planar views. Last but not least, we adopt an efficient dual-decomposition optimization approach for efficient inference of the 3D femur configuration from bi-planar views. Promising results demonstrate the potential of our method.

1 Introduction

Low-dose X-ray imaging has gained significant attention during the recent years, due to the increase of image quality and the decrease of the radiation exposure. Even though such images do not scale to computer tomography in terms of precision and resolution, they can be a valuable diagnostic tool for a number of diseases and in particular for osteopathies [10] like those related to spine [5], femur [9], *etc.* In this paper, we are interested in 3D proximal femur reconstruction focusing on highly accurate 3D modeling, in particular, for the femoral head part. In fact, patient specific 3D planning of the femoral head, inherits important diagnostic interest in related surgical interventions such as total hip replacement and intertrochanteric osteotomy [3].

Bone extraction and segmentation from bi-planar X-ray images is a challenging task due to the poor image quality, the fact that one has to compensate

* This work is supported by the European Research Council Starting Grant DIOCLES (ERC-STG-259112) and the MEDICEN Competitive Cluster sterEOS+ grant.

the partial support (because of the 3D-to-2D projection), *etc.* In such a context, conventional stereo-vision techniques [5] that recover 3D measurements from 2D projections do fail to provide clinically interesting results either because of the lack of correct correspondences between the two views or due to the sparsity of the obtained measurements. Thus, prior knowledge is usually considered, where statistical models of proximal femur is learned from a set of training examples [4]. Point-distribution models and in particular active shape models (ASMs) [1] have become a main stream and were also used for femur segmentation [4]. Pre-processing such as segmentation of femur contours [8] can facilitate the 3D femur pose estimation since the task is transformed into searching for a geometric mapping between the 3D model and the corresponding 2D multi-view silhouettes. Other methods act directly on the projection space and try to optimize global and local pose by measuring a projection energy, which can depend on key points [10], contours [9] or regional statistics [11]. However, these methods inherit various limitations. They need registration for all training examples to a common pose and subsequently for the testing image. Furthermore, important training set has to be considered towards capturing the variability of the femur, due to the global representation of shape models. Last, the inference can also be quite problematic due to the use of gradient-driven methods or the need of pre-segmentation, which is very challenging because of low signal-to-noise ratio.

In this paper, we propose a novel approach for femur estimation. Our contributions are threefold. First, we propose a non-uniform hierarchical decomposition of the shape prior of increasing clinical-relevant precision which is achieved through curvature driven unsupervised clustering acting on the geodesic distances between vertices. Second, we introduce a higher-order graphical-model representation of the femur which can be learned from a small number of training examples and involves third-order and fourth-order priors, while being similarity and mirror-symmetry invariant and providing means of measuring regional and boundary supports in the bi-planar views. Last but not least, we adopt an efficient dual-decomposition optimization approach for efficient inference of the 3D femur configuration from bi-planar views, leading to promising results.

The reminder of the paper is organized as follows: we present our hierarchical shape representation with the corresponding prior in Sec. 2, and a decomposed observation model in Sec. 3. They are combined within a probabilistic formulation which is then transformed into a higher-order MRF towards femur estimation in Sec. 4. Experimental results compose Sec. 5, while discussion and future work conclude the paper in Sec. 6.

2 Hierarchical Multi-resolution Probabilistic Modeling

In the literature, a surface is usually modeled as a mesh with uniform resolution. However, some anatomical regions (*e.g.*, femoral head) are of higher clinical relevance than adjacent parts. Hence, we propose to construct a hierarchical multi-resolution representation (Fig. 1(a)) of the femur with different clinical-relevant precisions, so that computational effort can be focused on the parts that are meant to be reconstructed with high fidelity.



Fig. 1. (a) Multi-resolution surface Model. (b) EOS System.

First, **mesh sub-sampling** from one level \mathcal{V}_m to a coarser one $\mathcal{V}_{m+1} \subset \mathcal{V}_m$ is performed iteratively, starting from a given high-resolution mesh $\mathcal{M}_0 = (\mathcal{V}_0, \mathcal{E}_0, \mathcal{F}_0)$ (\mathcal{V} denotes vertex set, \mathcal{E} edge set and \mathcal{F} face set), to create a non-increasing sequence of vertex subsets $\mathcal{V}_0 \supset \cdots \supset \mathcal{V}_M$, where each subset stands for a level of detail in the multi-resolution hierarchy. We propose to formulate such a sub-sampling problem as a clustering (Eq. [1](#)) based on geodesic distances and curvatures, aiming to obtain a subset \mathcal{V}_{m+1} of points uniformly distributed on the surface while preserving preferentially regions of high curvature:

$$\mathcal{V}_{m+1} = \operatorname{argmin}_{\mathcal{V} \subset \mathcal{V}_m} \left[\sum_{v \in \mathcal{V}_m} \min_{\hat{v} \in \mathcal{V}} d(v, \hat{v}) + \alpha \sum_{\hat{v} \in \mathcal{V}} \exp(-\operatorname{curv}(\hat{v})) \right] \quad (1)$$

where $d(v, \hat{v})$ is the geodesic distance between v and \hat{v} on \mathcal{M}_0 , α is a positive weight and $\operatorname{curv}(\hat{v})$ is the curvature at \hat{v} on \mathcal{M}_0 . The higher $\operatorname{curv}(\hat{v})$ is, the easier \hat{v} will be promoted as cluster center. Hence, this formulation tends to promote more cluster centers in the regions of high curvature. It can be efficiently solved through linear programming techniques proposed in [7](#) and the obtained \mathcal{V}_{m+1} corresponds to the cluster centers and each $v \in \mathcal{V}_m$ is associated to the closest center. One main particularity of this approach is that, as opposed to classical approaches (*e.g.*, K-means), the number of clusters is controlled by the penalty $\exp(-\operatorname{curv}(\hat{v}))$ which is related to flatness of the surface.

Second, **level of detail selection** is performed to select vertices from the sets $(\mathcal{V}_m)_{m=0, \dots, M}$ so that different regions are represented at given resolutions. We organize the vertices in a tree structure, since any vertex in \mathcal{V}_m is associated to one and only one vertex in \mathcal{V}_{m+1} in the hierarchical structure derived from the previous step. Then starting from the coarsest resolution, one can select the regions to be refined and iterate this process until reaching the required accuracy for every part. Through this step we obtain a set of vertices \mathcal{V}_{MR} .

Last, **connectivity computation** is performed on \mathcal{V}_{MR} in order to achieve the triangulated multi-resolution mesh. To this end, we compute the Delaunay triangulation of \mathcal{V}_{MR} associated to the geodesic distance. By viewing the Delaunay triangulation as the dual of the Voronoi diagram of \mathcal{V}_{MR} , it boils down to determining which pairs of Voronoi cells have a common boundary. In this way,

we obtain a set of edges \mathcal{E}_{MR} . Then the corresponding faces \mathcal{F}_{MR} are computed to achieve the mesh, by searching for minimal cycles in the edge list.

The proposed approach provides an economic way to model the surface according to the clinical relevance and a significantly better approximation of the original mesh compared with the classic edge collapse techniques. Furthermore, the vertices of the obtained mesh are among those of the original mesh, which facilitates largely the model prior learning. Finally, let $(\mathcal{V}, \mathcal{E}, \mathcal{F})$ denote the vertex set \mathcal{V}_{MR} , the edge set \mathcal{E}_{MR} and the face set \mathcal{F}_{MR} of the obtained mesh. Then the surface is parametrized by the 3D positions $\mathbf{u} = (u_i)_{i \in \mathcal{V}}$ of all the vertices.

Probabilistic Shape Modeling: The training data to learn statistical model consist of the 3D positions of the vertices on the training surfaces. No assumption on registration between surfaces is being made, we only assume that correspondences have been determined for the vertices among the training samples.

We adopt the pose-invariant prior in [12] and extend it in a more general formulation. Such a prior does not require the estimation of global pose in the training and testing stages, which eliminates the bias caused by such estimations. Furthermore, this prior is mirror-symmetry-invariant, leading to one more advantage in our problem: a common statistical model can be used for both the left and right femurs. For a clique c ($c \subseteq \mathbf{V}$ and $|c| \geq 3$) of vertices, we enumerate all the pairs $\mathcal{P}_c = \{(i, j) | i, j \in c \text{ and } i < j\}$ of points and compute the relative distance $\hat{d}_{ij} = d_{ij} / \sum_{(i,j) \in \mathcal{P}_c} d_{ij}$ for a pair of points $(i, j) \in \mathcal{P}_c$, where $d_{ij} = \|u_i - u_j\|$ denotes the Euclidean distance between points i and j . Since for clique c , any relative distance \hat{d}_{ij} is a linear combination of the others (*i.e.*, $\sum_{(i,j) \in \mathcal{P}_c} \hat{d}_{ij} = 1$), we put all the relative distances except one (denoted as \bar{p}) in a vector $\hat{\mathbf{d}}_c = (\hat{d}_{ij})_{(i,j) \in \mathcal{P}_c \setminus \{\bar{p}\}}$. We model the distribution $\psi_c(\hat{\mathbf{d}}_c)$ of $\hat{\mathbf{d}}_c$ using Gaussian Mixture Models (GMMs) which are learned from the training data.

In order to enforce the smoothness of the surface, for each quadruplet q of vertices corresponding to a pair of adjacent facets, we introduce a smoothness potential function encoding constraints on the change of the normal directions:

$$\psi_q(\mathbf{u}_q) = \exp \left\{ -\left(1 - \langle \vec{n}_q^{(1)}(\mathbf{u}_q), \vec{n}_q^{(2)}(\mathbf{u}_q) \rangle\right) / \beta \right\} \tag{2}$$

where \mathbf{u}_q denotes the 3D positions of the four vertices contained in q , $\langle \cdot, \cdot \rangle$ denotes the scalar product, β is a positive constant, and $\vec{n}_q^{(1)}, \vec{n}_q^{(2)}$ denote the unit normal vectors of the two facets.

Finally, let $\mathcal{T} = \{c | c \subseteq \mathcal{V} \text{ and } |c| = 3\}$ denotes the set of triplets¹, and $\mathcal{Q} \subset \{q | q \subseteq \mathcal{V} \text{ and } |q| = 4\}$ denote the set of quadruplets corresponding to all pairs of adjacent facets, then the prior probability of the 3D model is defined as:

$$p(\mathbf{u}) \propto \prod_{c \in \mathcal{T}} \psi_c(\hat{\mathbf{d}}_c(\mathbf{u}_c)) \cdot \prod_{q \in \mathcal{Q}} \psi_q(\mathbf{u}_q) \tag{3}$$

¹ We use 3-order cliques for the corresponding prior term, *i.e.*, $|c| = 3$. However, other higher-order cliques c ($|c| \geq 3$) can also be used.

3 Observation Model

Let $\mathbf{I} = (I_k)_{k \in \mathcal{K}}$ (where $\mathcal{K} = \{1, \dots, K\}$, $K = 2$ for the case of bi-planar views) denote K observed images captured from different viewpoints with the corresponding projection matrices $\mathbf{\Pi} = (\Pi_k)_{k \in \mathcal{K}}$. By exploiting the conditional independency, the image likelihood $p(\mathbf{I}|\mathbf{u}, \mathbf{\Pi})$ is decomposed into:

$$p(\mathbf{I}|\mathbf{u}, \mathbf{\Pi}) = \prod_{k \in \mathcal{K}} p(I_k|\mathbf{u}, \Pi_k) \quad (4)$$

We introduce a decomposed formulation for the likelihood $p(I_k|\mathbf{u}, \Pi_k)$ of the observed image k , which can be easily encoded in a higher-order MRF towards efficient inference. $p(I_k|\mathbf{u}, \Pi_k)$ is modeled as a Gibbs distribution, combining both the region-based and boundary-based data supports:

$$p(I_k|\mathbf{u}, \Pi_k) \propto \exp\left\{-\frac{\lambda E_k^{\text{R}}(I_k, \mathbf{u}, \Pi_k) + (1 - \lambda)E_k^{\text{B}}(I_k, \mathbf{u}, \Pi_k)}{T_k}\right\} \quad (5)$$

where T_k is temperature, and $0 < \lambda < 1$ is a balancing weight coefficient.

Regional Term. E_k^{R} encodes the statistical intensity properties of the two classes (femur and non-femur) and guides the projections of the surface to match the silhouettes of the femur in the observed images. We define this term as a sum of likelihoods on the projections of the front-facing facets:

$$E_k^{\text{R}}(I_k, \mathbf{u}, \Pi_k) = \sum_{f \in \mathcal{F}} \delta_f(\mathbf{u}_f, \Pi_k) \cdot \iint_{\Omega(\mathbf{u}_f, \Pi_k)} \log \frac{p_{bg}(I(x, y))}{p_{fg}(I(x, y))} dx dy \quad (6)$$

where f denotes a triangular facet, \mathbf{u}_f denotes the 3D coordinates of the three vertices of f , the indicator function $\delta_f(\mathbf{u}_f, \Pi_k)$ is equal to 1 if the facet f is front-facing w.r.t. the camera and 0 otherwise, $\Omega_f(\mathbf{u}_f, \Pi_k)$ is the 2D region corresponding to the projection of f , p_{fg} and p_{bg} denote the distributions (modeled using GMMs) of the intensity for the regions of the femur and the background.

Boundary Term. E_k^{B} encodes discontinuities along the boundaries. Recall that \mathcal{Q} denotes the set of quadruplets of vertices for the adjacent facets sharing an edge $e \in \mathcal{E}$. For each quadruplet $q \in \mathcal{Q}$, we define a function $\delta_q(\mathbf{u}_q, \Pi_k)$ which is equal to 1 (*i.e.*, the projection of e is a piece of boundary) when the facet closer to the camera is front-facing and the other one is back-facing, and 0 otherwise. We define the boundary term as a sum of the discontinuity measurement for all these quadruplets:

$$E_k^{\text{B}}(I_k, \mathbf{u}, \Pi_k) = \sum_{q \in \mathcal{Q}} \delta_q(\mathbf{u}_q, \Pi_k) \cdot \int_{\Gamma(\mathbf{u}_q, \Pi_k)} \langle \nabla I_k(x, y), \overrightarrow{n(x, y)} \rangle ds \quad (7)$$

where $\Gamma(\mathbf{u}_q, \Pi_k)$ denotes the projection of the edge shared by the two adjacent facets, $\overrightarrow{n(x, y)}$ denotes the outward-pointing unit normal of $\Gamma(\mathbf{u}_q, \Pi_k)$, $\nabla I_k(x, y) = (\frac{\partial I_k(x, y)}{\partial x}, \frac{\partial I_k(x, y)}{\partial y})$ denotes the gradient of the intensity at (x, y) . Note that other boundary discontinuity measurements can also be used as the integrand in Eq. [7](#).

4 Probabilistic 3D Surface Estimation Framework

The 3D model prior $p(\mathbf{u})$ (Sec. 2) and the image likelihood $p(\mathbf{I}|\mathbf{u}, \mathbf{\Pi})$ (Sec. 3) can be combined through a Bayesian framework, where the 3D surface estimation is formulated as a maximization of the posterior probability of \mathbf{u} :

$$\mathbf{u}^{\text{opt}} = \underset{\mathbf{u} \in \mathcal{U}}{\text{argmax}} p(\mathbf{u}|\mathbf{I}, \mathbf{\Pi}) \tag{8}$$

And the posterior probability $p(\mathbf{u}|\mathbf{I}, \mathbf{\Pi})$ is:

$$p(\mathbf{u}|\mathbf{I}, \mathbf{\Pi}) = \frac{p(\mathbf{u}, \mathbf{I}, \mathbf{\Pi})}{p(\mathbf{I}, \mathbf{\Pi})} \propto p(\mathbf{u}, \mathbf{I}, \mathbf{\Pi}) = p(\mathbf{I}|\mathbf{u}, \mathbf{\Pi})p(\mathbf{u})p(\mathbf{\Pi}) \propto p(\mathbf{I}|\mathbf{u}, \mathbf{\Pi})p(\mathbf{u}) \tag{9}$$

Higher-order MRF Formulation: The probabilistic framework above can be easily reformulated within a higher-order MRF so that we can employ efficient MRF inference algorithms to achieve femur reconstruction. To this end, we use a node to model a vertex i ($i \in \mathcal{V}$) with its latent 3D position U_i , a third-order clique c ($c \in \mathcal{T}$) to model a triplet of vertices, and a fourth-order clique q ($q \in \mathcal{Q}$) to model the quadruplets of vertices corresponding to the adjacent facets. The 3D model is estimated through minimizing the MRF energy $E(\mathbf{u})$:

$$\mathbf{u}^{\text{opt}} = \underset{\mathbf{u} \in \mathcal{U}}{\text{argmin}} E(\mathbf{u}) \tag{10}$$

where $E(\mathbf{u}) = -\log p(\mathbf{u}|\mathbf{I}, \mathbf{\Pi}) + \text{constant}$, and can be factorized into:

$$E(\mathbf{u}) = \sum_{f \in \mathcal{F}} H_f^R(\mathbf{u}_f) + \sum_{q \in \mathcal{Q}} (H_q^B(\mathbf{u}_q) + H_q^P(\mathbf{u}_q)) + \sum_{c \in \mathcal{T}} H_c^P(\mathbf{u}_c) \tag{11}$$

Regional-term potentials $H_f^R(\mathbf{u}_f)$ ($f \in \mathcal{F}$) encode the regional data term $E_k^R(I_k, \mathbf{u}, \Pi_k)$ (Eq. 6). Boundary-term potentials $H_q^B(\mathbf{u}_q)$ ($q \in \mathcal{Q}$) encode the boundary term $E_k^B(I_k, \mathbf{u}, \Pi_k)$ (Eq. 7). Model Prior Potentials $H_c^P(\mathbf{u}_c)$ ($c \in \mathcal{T}$) and $H_q^P(\mathbf{u}_q)$ ($q \in \mathcal{Q}$) encode the prior on the relative positions of points and the prior on the smoothness, respectively (Eq. 3). Through the negative logarithmic operation, we can obtain their definitions as follows:

$$\begin{cases} H_f^R(\mathbf{u}_f) = \sum_{k \in \mathcal{K}} \lambda_1^k \cdot \delta_f(\mathbf{u}_f, \Pi_k) \cdot \iint_{\Omega(\mathbf{u}_f, \Pi_k)} \log \frac{p_{bq}(I(x,y))}{p_{fg}(I(x,y))} dx dy \\ H_q^B(\mathbf{u}_q) = \sum_{k \in \mathcal{K}} \lambda_2^k \cdot \delta_q(\mathbf{u}_q, \Pi_k) \cdot \int_{\Gamma(\mathbf{u}_q, \Pi_k)} \langle \nabla I_k(x,y), \vec{n}(x,y) \rangle ds \\ H_c^P(\mathbf{u}_c) = -\log \psi_c(\hat{\mathbf{d}}_c(\mathbf{u}_c)) \\ H_q^P(\mathbf{u}_q) = \lambda_3 \cdot (1 - \langle \vec{n}_q^{(1)}(\mathbf{u}_q), \vec{n}_q^{(2)}(\mathbf{u}_q) \rangle) \end{cases} \tag{12}$$

where λ_1^k , λ_2^k and λ_3 are positive constant weights² for the corresponding terms.

Dual-Decomposition MRF Inference. [6] was adopted to optimize the proposed higher-order MRF. Like [12], we decompose the original graph into a set of factor trees which can be solved within polynomial time using max-product belief propagation. Their solutions are combined using projected subgradient method [6] to achieve the solution of the original problem.

² We use a single constant λ_1 for λ_1^k ($\forall k$), and a single constant λ_2 for λ_2^k ($\forall k$).

5 Experimental Validation

We use the low-dose EOS imaging system (BiospaceMedTM) which is an alternative system that captures simultaneously bi-planar X-ray images, coronal and sagittal in an upright position (Fig. 1(b)). The training set consists of 20 CT patient volumes of the hip. Manual segmentations of them are provided to extract femoral surfaces, which are used to learn the model prior. With the obtained prior, we perform the femur reconstruction using the proposed method. The search of the model parameters is done using the coarse-to-fine scheme and the sparse sampling strategy as in 2.

The proposed method has been validated using both dry femurs and real clinical data. The in vitro testing data consist of 14 pairs of EOS images, for which the ground truth surfaces are provided by the gold standard CT method. The bones are of real size for adults. Quantitative evaluation was conducted by comparing the estimated surfaces to the ground truths, with respect to the DICE coefficient and the distances from the 3D vertices of the estimated femur to the ground truth surface (called point-to-surface error). Fig. 2(a) shows four samples from the set of results, with point-to-surface error on femoral head. Fig. 2(b) presents the statistics of the DICE coefficients, the mean and standard deviation of the point-to-surface errors for the whole in vitro validation set. Four in vivo examples were also tested, where both left and right femurs were constructed from the EOS images. The qualitative results in Fig. 2(c-d) show that the projections of the reconstructed surfaces fit well the femurs in the observed images despite complex scene and low image quality.

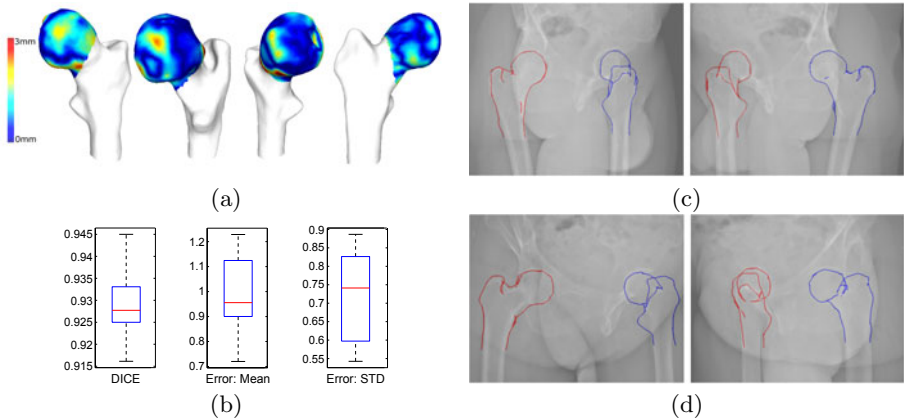


Fig. 2. Experimental Results. (a) Four 3D surface reconstruction results with point-to-surface errors on femoral head. (b) Boxplots on the DICE, the mean and STD of the point-to-surface errors (mm). On each box, the central mark in red is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points. (c) and (d) Results on in vivo data, where projected boundaries of the estimated femurs in the bi-planar views are shown for each example.

6 Conclusion

In this paper, we have proposed a novel approach to 3D femur estimation from bi-planar X-ray views. The main innovations consist in the multi-resolution shape representation, the ability to model pose/mirror-symmetry invariant prior and the image likelihood through local interactions, and the higher-order MRF formulation of the surface estimation which is solved by dual-decomposition optimization. Promising results demonstrate the potential of the method.

Future work consists of introducing a joint model that couples femur with the hipbone socket which could enhance the diagnostic potential of the method, and combining distinctive (visual) anatomical landmarks with the existing formulation towards increasing precision and the overall performance. Last, the application of the method to other clinical settings also bears great promises.

References

1. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active shape models - their training and application. *CVIU* 61(1), 38–59 (1995)
2. Glocker, B., Komodakis, N., Tziritas, G., Navab, N., Paragios, N.: Dense image registration through mrfs and efficient linear programming. *Medical Image Analysis* 12(6), 731–741 (2008)
3. Gottschling, H., Roth, M., Schweikard, A., Burgkart, R.: Intraoperative, fluoroscopy-based planning for complex osteotomies of the proximal femur. *Int. J. Med. Robotics Comput. Assist. Surg.* 1, 67–73 (2005)
4. Heimann, T., Meinzer, H.P.: Statistical shape models for 3d medical image segmentation: A review. *Medical Image Analysis* 13(4), 543–563 (2009)
5. Kadoury, S., Cheriet, F., Laporte, C., Labelle, H.: A versatile 3d reconstruction system of the spine and pelvis for clinical assessment of spinal deformities. *Med. Biol. Engineering and Computing* 45(6), 591–602 (2007)
6. Komodakis, N., Paragios, N., Tziritas, G.: MRF optimization via dual decomposition: Message-passing revisited. In: *ICCV* (2007)
7. Komodakis, N., Paragios, N., Tziritas, G.: Clustering via lp-based stabilities. In: *NIPS* (2009)
8. Kurazume, R., Nakamura, K., Okada, T., Sato, Y., Sugano, N., Koyama, T., Iwashita, Y., Hasegawa, T.: 3d reconstruction of a femoral shape using a parametric model and two 2d fluoroscopic images. *CVIU* 113(2), 202–211 (2009)
9. Laporte, S., Skalli, W., de Guise, J.A., Lavaste, F., Mitton, D.: A biplanar reconstruction method based on 2d and 3d contours: Application to the distal femur. *Computer Methods in Biomechanics and Biomedical Engineering* 6(1), 1–6 (2003)
10. Mitton, D., Deschenes, S., Laporte, S., Godbout, B., Bertrand, S., de Guise, J.A., Skalli, W.: 3d reconstruction of the pelvis from bi-planar radiography. *Computer Methods in Biomechanics and Biomedical Engineering* 9(1), 1–5 (2006)
11. Varshney, K.R., Paragios, N., Deux, J.F., Kulski, A., Raymond, R., Hernigou, P., Rahmouni, A.: Postarthroplasty examination using x-ray images. *IEEE Trans. Med. Imaging* 28(3), 469–474 (2009)
12. Wang, C., Teboul, O., Michel, F., Essafi, S., Paragios, N.: 3d knowledge-based segmentation using pose-invariant higher-order graphs. In: Jiang, T., Navab, N., Pluim, J.P., Viergever, M.A. (eds.) *MICCAI 2010, Part III. LNCS, vol. 6363*, pp. 189–196. Springer, Heidelberg (2010)

Fully Automatic Segmentation of Brain Tumor Images Using Support Vector Machine Classification in Combination with Hierarchical Conditional Random Field Regularization

Stefan Bauer*, Lutz-P. Nolte, and Mauricio Reyes

Institute for Surgical Technology and Biomechanics,
University of Bern, Switzerland
stefan.bauer@istb.unibe.ch

Abstract. Delineating brain tumor boundaries from magnetic resonance images is an essential task for the analysis of brain cancer. We propose a fully automatic method for brain tissue segmentation, which combines Support Vector Machine classification using multispectral intensities and textures with subsequent hierarchical regularization based on Conditional Random Fields. The CRF regularization introduces spatial constraints to the powerful SVM classification, which assumes voxels to be independent from their neighbors. The approach first separates healthy and tumor tissue before both regions are subclassified into cerebrospinal fluid, white matter, gray matter and necrotic, active, edema region respectively in a novel hierarchical way. The hierarchical approach adds robustness and speed by allowing to apply different levels of regularization at different stages. The method is fast and tailored to standard clinical acquisition protocols. It was assessed on 10 multispectral patient datasets with results outperforming previous methods in terms of segmentation detail and computation times.

Keywords: MRI, Segmentation, Brain Tumor, Glioma, SVM, CRF.

1 Introduction

Segmentation of healthy and pathologic brain tissues from magnetic resonance images (MRI), including their subregions, is important in cancer treatment planning as well as for cancer research. In current clinical practice, the analysis of brain tumor images is mostly done manually. Apart from being time-consuming, this has the additional drawback of significant intra- and interrater variability, which was reported to be around $20\% \pm 15\%$ and $28\% \pm 12\%$ respectively according to Mazzara et al. [8].

Nowadays, in clinical practice, usually four different MRI modalities are used to delineate the tumor and its subregions from MRI images of the head. These

* Funding by the European Union within the framework of the ContraCancrum project (FP7 - IST-223979) is gratefully acknowledged.

modalities include T_1 -weighted images (T_1), T_1 -weighted images with contrast enhancement (T_{1c}), T_2 -weighted images (T_2) and $T_{2\text{-flair}}$ -weighted images (T_{2f}). The tumor area is usually divided into necrotic, active and edema subregion. Each of the modalities reveal different subregions. In general, the radiologist considers all these MRI modalities simultaneously when segmenting brain tumor images. Standard clinical acquisition protocols for brain tumor imaging provide high intra-slice resolution, but poor inter-slice spacing. This poses an additional challenge for automatic analysis. In order to cope with all these challenges, we developed a new method, which is clinically-oriented and does not only make use of sophisticated image-processing methods, but also takes the type of data into account, which is usually available for large-scale clinical studies and patient treatment planning.

In this study, we focus on the case of gliomas, which is the most aggressive type of brain tumors. We treat the problem at hand as a classification task that maps each voxel to its corresponding label based on a multidimensional feature vector. Machine learning techniques have proven to yield good results in many cases. For example Verma et al. [11] use SVMs to classify brain tumors from a large number of different MRI modalities. However, these methods assume that the data is independent and identically distributed (iid), which is clearly not the case for image voxels. Most voxel labels strongly depend on their neighbors. These spatial relationships can be effectively considered using Conditional Random Field (CRF) methods. We suggest to use the CRF method to regularize the result produced by the classifier retrospectively, in a way inspired by the approach of Lee et al. [6]. The classifier is first trained under the assumption that the data is iid and the result is regularized subsequently using a CRF approach. This makes the whole process very efficient while maintaining improved accuracy by considering neighbor relationships.

In order to exploit the full potential of the classification method, we propose not only to segment the brain into healthy and tumor regions as it is usually done, but also subdivide the healthy region into cerebrospinal fluid (CSF), gray matter (GM) and white matter (WM) and subdivide the tumor area into necrotic part, active part and edema part. We suggest to employ a novel hierarchical approach for this task in order to increase robustness and speed. The hierarchical approach allows us to apply a strong regularization after segmenting the brain into tumor and healthy regions. Each of the two initially identified regions is then classified into its respective subregions before a weaker regularization is applied to the labeled image of the healthy and tumor subregions.

2 Methods

2.1 Feature Extraction

Features are extracted from the multispectral imaging data. The most prominent features for distinguishing pathologic and healthy tissues, as well as all their subregions, are the image intensities in the different modalities. Additionally, we use the first order texture features (mean, variance, skewness, kurtosis, energy,

entropy) according to [10]. First-order textures can be computed fast and easily from small patches around each voxel in all four modalities. This yields a 28-dimensional feature vector \mathbf{x} , which consists of the voxel-wise concatenation of the multimodal intensities I and multimodal textures \mathbf{T} at each voxel i as shown in equation (II).

$$\mathbf{x}(i) = [I_{T_1}(i), I_{T_{1c}}(i), I_{T_2}(i), I_{T_{2f}}(i), \mathbf{T}_{T_1}(i), \mathbf{T}_{T_{1c}}(i), \mathbf{T}_{T_2}(i), \mathbf{T}_{T_{2f}}(i)] \quad (1)$$

2.2 Classification

Classification is done using a soft-margin SVM classifier [9]. SVMs are discriminative classifiers, originating from machine learning. They require a training step to find a separating hyperplane for the data in the feature space. SVMs solve the optimization problem

$$\min_{\mathbf{w}, b, \xi} \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^l \xi_i \quad (2)$$

$$\text{subject to } y_i(\mathbf{w}^T \phi(\mathbf{x}_i) + b) \geq 1 - \xi_i, \text{ with } \xi_i \geq 0 \quad (3)$$

where (\mathbf{x}_i, y_i) are the instance-label pairs of the dataset and \mathbf{w} is the normal vector of the separating hyperplane. C is a penalty parameter for the error term and b is the offset of the hyperplane. The appealing property of SVMs is that they offer the possibility to use a kernel function $K(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j)$ for transforming the data into a higher-dimensional feature space, where the data can be linearly separated efficiently with a maximum margin. Slack variables ξ_i are used for soft-margin classification. Parameter selection for the SVM classifier with a radial basis function (RBF) kernel is done using grid-based cross-validation on the training data. In order to extend the inherently binary SVM classifier to a multiclass problem, we use a one-against-one voting strategy. Our SVM classification is based on the LibSVM implementation described in [2].

2.3 Regularization

Regularization is done in two different stages using a CRF method. Conditional Random Fields are an extension of Markov Random Fields (MRF). Both offer the possibility to formulate the regularization as an energy minimization problem. To this end we use a second-order CRF with two energy terms.

$$E = \sum_i V(y_i) + \sum_{ij} w(y_i, y_j, \mathbf{x}_i, \mathbf{x}_j) D_{pq}(y_i, y_j) \quad (4)$$

The first term in equation (4) denotes the data energy, which is only dependent on the data at the current point, whereas the second term constitutes the smoothness energy, which takes the neighborhood information into account. For the case at hand, V is the unary potential function, w describes the neighborhood relationships and D_{pq} is a label distance function for the pairwise potentials. The unary potentials can be calculated directly from the voxel-wise output produced

by the SVM classifier. The smoothness energy is computed depending on the neighboring voxels.

We compute the unary potentials $V(y_i)$ from the label output y_i of the SVM classifier. In equation (5), c_1 is a constant which allows us to adjust the weights of unary and pairwise potentials. \tilde{y}_i is the newly assigned label after regularization and δ is the Kronecker δ function.

$$V(y_i) = c_1 \cdot (1 - \delta(\tilde{y}_i, y_i)) \quad (5)$$

For computing the pairwise potentials, we suggest a new formulation, which is an extension of the approach given by Boykov et al. [1]. In equation (6), $c_2(i, j)$ is a weighting function, which allows us to apply different weights for different neighbors. The term $(1 - \delta(y_i, y_j))$ is the most important, penalizing different labels y of adjacent voxels. The last term in equation (6) imposes strong smoothness constraints in regions of similar intensities \mathbf{x} and relaxes regularization in regions of high intensity contrast. The intensity term is adapted for multimodal images and \bar{x} is a generalized multimodal mean intensity. The label distance function D_{pq} in equation (4) is used to penalize adjacencies of necrotic or active tumor regions with healthy tissues more strongly because these adjacencies are less likely to occur.

$$w(y_i, y_j, \mathbf{x}_i, \mathbf{x}_j) = c_2(i, j) \cdot (1 - \delta(y_i, y_j)) \cdot \exp\left(\frac{\|\mathbf{x}_i - \mathbf{x}_j\|}{2 \cdot \bar{x}}\right) \quad (6)$$

As mentioned before, the regularization is carried out in a hierarchical way. After the first coarse classification into tumor and healthy tissues, a strong 3D regularization is employed using a von Neumann neighborhood (6-neighboring voxels in 3D). However, in our approach costs w_{ij} in z-direction are weighted with a smaller constant $c_2(i, j)$ because the z-spacing of slices is much larger than the in-slice-spacing. If the same weight would be applied to all neighbors, the label map would be oversmoothed in z-direction.

In a second stage, regularization is applied on the image, which has been subclassified into the different tumor and healthy subregions. In this case very fine structures, especially for CSF, are present. These fine structures do not allow to use a 3D regularization. Due to the large z-spacing, the fine structures would be oversmoothed. Therefore, we use only a 2D neighborhood for the second stage, but with a Moore neighborhood structure now (8 neighbors in 2D), which renders smoother results.

Optimizing CRFs is a challenging and computationally intensive task, especially when many voxels and multiple labels are involved. We employ a recent optimization algorithm proposed by Komodakis et al. [5], which is based on linear programming via graph-cuts and primal-dual decomposition. The algorithm is able to efficiently derive almost optimal solutions for a wide class of NP-hard CRF problems. The most important reason for us to use this optimization approach is its great computational efficiency, which renders it much faster than conventional graph-cuts optimization techniques.

2.4 Application to Brain Tumor Image Analysis

Initially, the images undergo a preprocessing pipeline. In a first step, the four modalities are registered with the help of a rigid registration and mutual information metric. Next, only the brain region is extracted from the images using a fully-automatic, customized skull-stripping algorithm^[7]. Subsequently, noise is removed with an edge-preserving smoothing filter and the bias field is corrected. In order to allow for interpatient classification, the image histograms are matched across all patients for each MRI modality separately. The preprocessing is completely integrated with the SVM classification and CRF regularization components using the Insight Toolkit for Segmentation and Registration (ITK)^[4].

3 Results

3.1 Image Data

We evaluated our algorithm on images of 10 patients from the ContraCancrum brain tumor database^[7]. Each patient dataset consists of a T_1 , T_{1c} , T_2 and T_{2f} image. The images were resampled to an in-slice resolution of 1mm (210x210 to 260x260 voxels), the inter-slice spacing was between 3mm and 6mm depending on the dataset (19 to 25 slices). Figure^[1] shows an axial slice of one patient with the four MRI modalities we used.

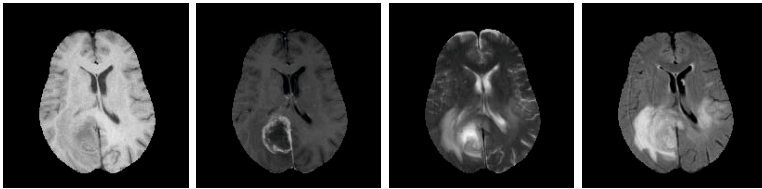


Fig. 1. Axial slice of one patient. The four MRI modalities under study are shown, from left to right: the T_1 , T_{1c} , T_2 and T_{2f} image.

3.2 Evaluation

For quantitative evaluation of the results we use the Dice similarity coefficient. The ground truth was defined by a manual segmentation. However, the manual segmentation was only available for the tumor tissues, not for the healthy tissues. Therefore only the accuracy of the tumor segmentation could be evaluated quantitatively. The accuracy of the segmentation of healthy tissues could only be rated qualitatively by visual inspection.

We subdivided our results into intra- and interpatient case. For the intrapatient case, the classifier was trained on small subregions of the relevant areas

¹ available from:

http://www.istb.unibe.ch/content/surgical_technologies/medical_image_analysis/software/

in the same patient, while in the interpatient case we performed leave-one-out cross-validation. Additionally, in both cases, we compared our proposed hierarchical, regularized approach with a non-hierarchical classification, that does not comprise the proposed two-step CRF-regularization.

Figure 2 presents the segmentation result in case of interpatient training using leave-one-out cross-validation for the same slice as shown in figure 1. The gross tumor volume (GTV), comprising all tumor tissues, is well delineated, although the segmentation appears noisy if no regularization is applied. Classifying the tumor subregions appears to be more challenging, but our proposed approach performs still better than the method without hierarchy and regularization. From a visual inspection, the segmentation of healthy tissues (CSF,

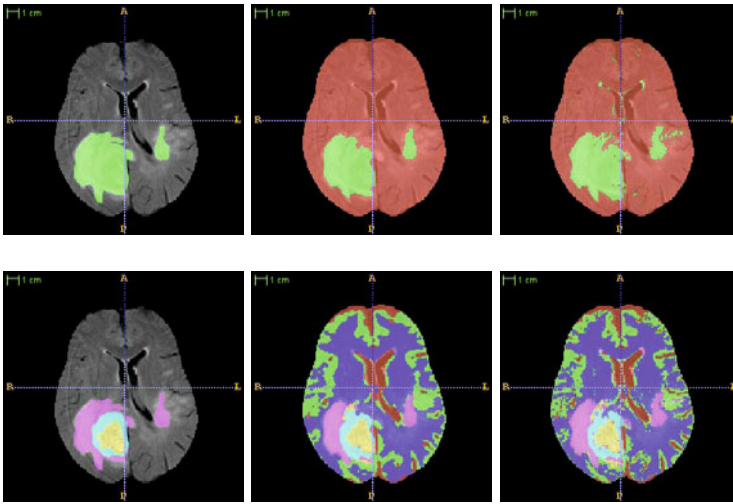


Fig. 2. Results for one axial slice of the interpatient leave-one-out case. First row: coarse classification into tumor and healthy tissues, second row: fine classification into necrotic, active, edema part and CSF, GM, WM respectively. From left to right: manual segmentation, hierarchical SVM-classification with CRF-regularization, non-hierarchical SVM-classification without regularization.

Table 1. Dice similarity coefficient (mean and standard deviation) for the ten datasets under study. Gross tumor volume (GTV) comprises the complete tumor region, including active, necrotic and edema part. Results are shown for the intra- and interpatient (leave-one-out cross-validation) case, with and without regularization.

	GTV	Necrotic	Active	Edema
Inpatient Regularized	0.84 ± 0.03	0.61 ± 0.24	0.71 ± 0.09	0.73 ± 0.04
Inpatient Unregularized	0.76 ± 0.10	0.45 ± 0.31	0.59 ± 0.16	0.72 ± 0.07
Interpatient Regularized	0.77 ± 0.09	0.45 ± 0.23	0.64 ± 0.13	0.60 ± 0.16
Interpatient Unregularized	0.67 ± 0.13	0.30 ± 0.24	0.46 ± 0.12	0.63 ± 0.16

GM, WM) appears to be reasonable in all cases, but again the result has many outliers if no regularization is applied.

The quantitative results using Dice similarity coefficient are summarized in table 1. With hierarchical regularization, the mean Dice similarity coefficient over all 10 patients is 0.84 for the inpatient case and 0.77 for the interpatient leave-one-out case when the GTV is considered. Dice coefficients for the individual tumor subregions are lower. However, the results show once again that the Dice coefficient is clearly worse when no hierarchical regularization is applied.

Computation time for the segmentation algorithm (excluding preprocessing) is between 20 and 120 seconds on a single CPU running at 2.33 GHz. Computation time mainly depends on the size of the image dataset and on the complexity of the SVM optimization.

4 Discussion and Conclusion

We presented a clinically-oriented method to segment 3D MRI images of brain tumor patients into healthy and tumor areas, including their individual subregions. For this, we propose to apply a hierarchical SVM-based classification and combine it with a CRF-based regularization in two stages.

Dice similarity coefficients for the GTV are in a similar range as the results presented in [3] and [12], but different data was used. In contrast to those approaches, our method additionally returns all the tumor and healthy subregions while being faster in computation time. The accuracy of our automatic method lies in a similar range as the values reported for inter-observer variability of manual segmentations in [8]. However, the automatic method has advantages in longitudinal studies because the results are not biased subjectively. Due to the additional difficulty in subdividing the tumor region, Dice similarity coefficients for the individual tumor subregions are lower than for the gross tumor volume.

We show that the combination of SVM and CRF methods offers two complementary strengths. The results mentioned in table 1 suggest that our approach including hierarchical regularization improves the overlap with the ground truth significantly. This regularization seems to be even more important when training is applied on different patients. With the hierarchical regularization, the Dice coefficients are only slightly worse in interpatient training compared to inpatient training. This is an important finding because the possibility to perform training on a disjoint subset of patients is crucial for efficient segmentation procedures in clinical practice. A visual inspection of the images in figure 2 shows that the segmentation results for healthy tissues as well as for tumor tissues appear much noisier when no CRF regularization is applied and inconsistencies of adjacent slices can be observed.

In contrast to other methods, our proposed segmentation approach yields the tumor and healthy region, as well as all their subregions. It adds powerful spatial regularization to iid classification and runs fast on clinically relevant data.

References

1. Boykov, Y., Jolly, M.: Interactive graph cuts for optimal boundary & region segmentation of objects in ND images. In: Proc. ICCV, pp. 105–112. IEEE, Los Alamitos (2002)
2. Chang, C.C., Lin, C.J.: LIBSVM: a library for support vector machines (2001), <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
3. Corso, J.J., Sharon, E., Dube, S., El-Saden, S., Sinha, U., Yuille, A.: Efficient multilevel brain tumor segmentation with integrated bayesian model classification. *IEEE T. Med. Imaging* 27(5), 629–640 (2008)
4. Ibanez, L., Schroeder, W., Ng, L., Cates, J.: The ITK Software Guide. Kitware, 2nd edn. (2005), <http://www.itk.org/ItkSoftwareGuide.pdf>
5. Komodakis, N., Tziritas, G., Paragios, N.: gle and dynamic MRFs: Setting the state of the art with primal-dual strategies. *Comput. Vis. Image Und.* 112(1), 14–29 (2008)
6. Lee, C., Wang, S., Murtha, A., Brown, M., Greiner, R.: Segmenting Brain Tumors using Pseudo-Conditional Random Fields. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 359–366. Springer, Heidelberg (2008)
7. Marias, K., Dionysiou, D., Sakkalis, V., Graf, N., Bohle, R.M., Coveney, P.V., Wan, S., Folarin, A., Buechler, P., Reyes, M., Clapworthy, G., Liu, E., Sabczynski, J., Bily, T., Roniotis, A., Tsiknakis, M., Giatili, S., Veith, C., Messe, E., Stenzhorn, H., Kim, Y.j., Zasada, S., Haidar, A.N., Bauer, S., Wang, T., Zhao, Y., Karasek, M., Grewer, R., Franz, A., Stamatakos, G.: Clinically driven design of multi-scale cancer models: the ContraCancrum project paradigm. *J. Roy. Soc. Interface Focus* 1(3), 450–461 (2011)
8. Mazzara, G., Velthuizen, R., Pearlman, J., Greenberg, H., Wagner, H.: Brain tumor target volume determination for radiation treatment planning through automated MRI segmentation. *Int. J. Radiat. Oncol., Biol. Phys.* 59(1), 300–312 (2004)
9. Schoelkopf, B., Smola, A.: Learning with kernels: support vector machines, regularization, optimization, and beyond. MIT Press, Cambridge (2002)
10. Tuceryan, M., Jain, A.: Texture analysis, 2nd edn., pp. 207–248. World Scientific Publishing, Singapore (1998)
11. Verma, R., Zacharaki, E., Ou, Y., Cai, H., Chawla, S., Lee, S., Melhem, E., Wolf, R., Davatzikos, C.: Multiparametric Tissue Characterization of Brain Neoplasms and Their Recurrence Using Pattern Classification of MR Images. *Acad. Radiol.* 15(8), 966–977 (2008)
12. Wels, M., Carneiro, G., Aplas, A., Huber, M., Hornegger, J., Comaniciu, D.: A Discriminative Model-Constrained Graph Cuts Approach to Fully Automated Pediatric Brain Tumor Segmentation in 3-D MRI. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 67–75. Springer, Heidelberg (2008)

Midbrain Segmentation in Transcranial 3D Ultrasound for Parkinson Diagnosis

Seyed-Ahmad Ahmadi¹, Maximilian Baust¹, Athanasios Karamalis¹,
Annika Plate², Kai Boetzel², Tassilo Klein¹, and Nassir Navab¹

¹ Computer Aided Medical Procedures, Technische Universität München, Germany
ahmadi@cs.tum.edu

² Department of Neurology, Ludwig-Maximilians-University of Munich, Germany

Abstract. Ultrasound examination of the human brain through the temporal bone window, also called transcranial ultrasound (TC-US), is a completely non-invasive and cost-efficient technique, which has established itself for differential diagnosis of Parkinson’s Disease (PD) in the past decade. The method requires spatial analysis of ultrasound hyperechogenicities produced by pathological changes within the Substantia Nigra (SN), which belongs to the basal ganglia within the midbrain. Related work on computer aided PD diagnosis shows the urgent need for an accurate and robust segmentation of the midbrain from 3D TC-US, which is an extremely difficult task due to poor image quality of TC-US. In contrast to 2D segmentations within earlier approaches, we develop the first method for semi-automatic midbrain segmentation from 3D TC-US and demonstrate its potential benefit on a database of 11 diagnosed Parkinson patients and 11 healthy controls.

1 Introduction and Medical Motivation

The midbrain, also called mesencephalon, is an approximately 2x2x1cm sized region of the brain, which forms the upper part of the brainstem. It contains several conglomerates of nerve cells, so called nuclei, which are involved in visual, auditory, and motoric functions of the human brain. It is well known that a degeneration of the *Substantia nigra* (SN), cf. Fig. 1, which is located in the midbrain, is the cause of Parkinson’s Disease (PD) [16]. This degeneration, which is accompanied by an increased concentration of ferrite deposits, depletes other parts of the brain from the neurotransmitter dopamine, thereby causing the cardinal motor symptoms of Parkinson’s disease, i.e. slowed movement, tremor, and muscular rigidity. These ferrite deposits can be visualized by transcranial B-mode ultrasound (TC-US) in form of hyperechogenic areas within the SN [1].

Several studies using this completely non-invasive method [1][16][2] found that it is possible to distinguish PD patients and PD types with sensitivities and specificities of above 90%, based on the size of the hyperechogenic area within the SN. There is also the indication that SN deterioration is visible in US *before* onset of motor symptoms, which is of high medical relevance, because an early diagnosis of PD may lead to better treatment and deceleration of disease

progression. Consequently, the hope of medical researchers is to be able to use TC-US as a screening method for PD in future, especially since the degeneration of SN cells *cannot* be visualized by Magnetic Resonance Imaging (MRI) [12].

Though bearing a lot of potential, PD diagnosis based on transcranial US is not easy and suffers from a few issues. Due to the scanning at low frequencies (2-4 MHz), transcranial US is unfortunately poor in resolution and characterized by high levels of noise and large speckle patterns (see Fig. 1). Also, the US acquisition through the temporal bone window, a layer of bone thin enough to be penetrated by low-frequency US (typically 2-4MHz), introduces additional phase aberrations and interferences [9] and higher absorption than in regular tissue, causing low contrast. Moreover, a high amount of experience is necessary to find an optimal scanplane in 2D and to interpret TC-US images correctly [15]. Thus, PD diagnosis based on TC-US suffers from high intra- and inter-rater variability. We strongly believe that this procedure can be improved significantly by the usage of 3D-US, making the selection of a cross-section superfluous, as well as an automated processing of the volumetric data, making the segmentation results reproducible.

The crucial part of any computer aided diagnosis system is an accurate segmentation of the whole midbrain. In fact, it is advisable to segment the midbrain first and the SN afterwards, because SN echogenicities appear to be almost identical to speckle and intensity patterns outside the midbrain area (cf. US images in Fig. 1). This was also proven by related work on computer aided PD diagnosis, where either a manual [10,4] or an automatic [6] midbrain segmentation is used.

As a consequence, any method which provides a robust and reliable midbrain segmentation from transcranial US greatly facilitates the subsequent steps of SN segmentation and classification. Please note that the only related approach proving an automated midbrain segmentation is the one of [6]. This method is, however, restricted to 2D and based on a complex and computationally intensive finite-element model. We argue that no further attempts at computer-assisted segmentation of transcranial B-Mode US have been performed yet, because a careful choice of prior knowledge, data term and optimization method is necessary (see Results section). Indeed, even recent reviews, e.g. [8], show that the application of shape models to 3D-US segmentation has not been addressed extensively, so far. To the best of our knowledge, this is the first work on *volumetric* midbrain segmentation from transcranial 3D-US.

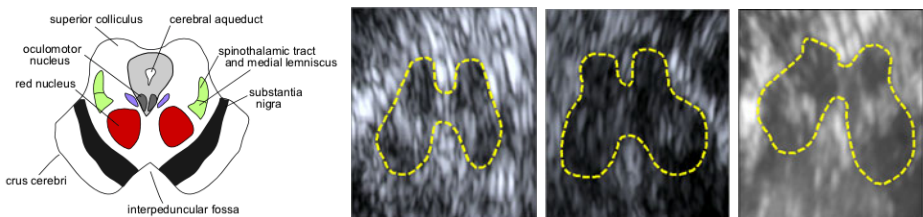


Fig. 1. From left to right: Midbrain and basal ganglia such as substantia nigra, along with typical 2D TC-US images (middle), and a slice through one of our 3D volumes

2 Materials and Methods

In the following, we will show that it is possible to construct a method which is able to provide a reliable segmentation of the midbrain in the area around the SN. The proposed approach is based on three components: Firstly, the generation of a statistical shape model. Secondly, the combination of this shape model with an active surface framework. Thirdly, the active polyhedron framework of [14] to implement the discrete surface evolution. Before we describe these three steps, we briefly describe the acquired data and the necessary pre-processing.

Data Acquisition and Pre-processing. For our study, we collected data from 23 subjects, 11 previously diagnosed PD patients and 11 healthy controls (mean age 60.7 yrs., 60% female, 40% male). One subject was excluded due to insufficient bone window (4.3% exclusion rate). A medical expert scanned the 22 subjects with a medical ultrasound machine¹ at 3 MHz. Scans were performed bi-laterally, i.e. through the left and right temporal bone window, taking care that the entire midbrain area was contained in the scanned area. Bi-lateral 2D US planes were reconstructed into a 3D Freehand US volume using a backward-warping compounding technique [17] at an isotropic resolution of 0.45mm. Afterwards, the medical expert segmented all 22 volumes in a blinded, anonymized fashion into 3 regions, midbrain, SN left and SN right. Segmentation was performed only for those slices in which the SN was clearly visible and considered relevant for a diagnostic decision. Manual midbrain segmentations were then transformed into 3D meshes using the MATLAB iso2mesh toolbox [7] and default parameters (i.e. head surface element size bound of 2, keep ratio of 0.05).

Statistical Shape Model Creation. For SSM creation, we utilize the comprehensive SPHARM-MAT toolbox of [13]. The toolbox was slightly modified to take a set of training meshes as input and generate the SSM model in one function. The only modified parameter was the SPHARM degree for expansion and it was set to 15 for all meshes, otherwise default parameters as downloaded with the toolbox were used. A 5-fold cross-validation was performed, i.e. the 22 meshes were split into five groups and five SSM models were generated from four groups (training set) to perform segmentation and evaluation on the fifth group (test set). The output of the SPHARM toolbox trained on $M + 1$ training shapes is a SSM with M modes of variation \mathcal{S}_i , $i = 1 \dots M$:

$$\mathcal{S}_\alpha = \mathcal{S}_\mu + \sum_{i=1}^M \alpha_i \mathcal{S}_i, \quad (1)$$

where \mathcal{S}_μ denotes the mean shape and $\alpha_1, \dots, \alpha_M \in \mathbb{R}$. The shape model \mathcal{S}_α can be considered as a parametrized family of shapes, where a specific configuration is completely determined by the shape vector $\alpha = (\alpha_1, \dots, \alpha_M)^T$.

¹ Siemens Accuson Antares, PX4-1 transducer.

Evolving the Shape Model. In order to evolve the shape model towards the desired configuration, we iteratively minimize an active surface energy of the form

$$E(\mathcal{S}) = \int_{int\mathcal{S}} f_i dx + \int_{ext\mathcal{S}} f_e dx, \tag{2}$$

where \mathcal{S} denotes the surface, $int\mathcal{S}$ the region inside \mathcal{S} , and $ext\mathcal{S}$ the region outside \mathcal{S} . As our shape model provides enough regularity itself we employ no additional regularizer, such as the surface area. Due to the highly inhomogeneous nature of US images, foreground and background regions cannot be described by global statistics. Thus, we use a localized version of the Chan-Vese model [3] proposed by [11]

$$\begin{cases} f_i = (I - c_i(x))^2, & \text{where } c_i(x) = \frac{\int_{int\mathcal{S}} B_\epsilon(x) I(x) dx}{\int_{int\mathcal{S}} B_\epsilon(x) dx}, \\ f_e = (I - c_e(x))^2, & \text{where } c_e(x) = \frac{\int_{ext\mathcal{S}} B_\epsilon(x) I(x) dx}{\int_{ext\mathcal{S}} B_\epsilon(x) dx}, \end{cases} \tag{3}$$

where $B_\epsilon(x)$ denotes a ball of radius ϵ centered at x . In order to derive the evolution equation for the shape model, i.e. the gradient descent for the shape vector α , we plug the shape model (1) into (2) and compute the partial derivatives with respect to α_j :

$$\frac{\partial}{\partial \alpha_j} E(\mathcal{S}_\alpha) = \frac{\partial E}{\partial \mathcal{S}} \frac{\partial \mathcal{S}}{\partial \alpha_j} = \int_{\mathcal{S}} (f_i - f_e) N \cdot \mathcal{S}_j ds, \tag{4}$$

where N denotes the surface normal. Put together, these partial derivatives yield the gradient of E with respect to the shape vector α , which we denote by $\nabla_\alpha E$. In order to implement a gradient descent for α , we have to discretize the expression in (4).

Active Polyhedron Framework. We decided to use an explicit surface representation based on a triangular mesh as it is given by the active polyhedron method described in [14]. However, as the created shape model provides enough regularity we do not use the regularization described in [14] and evolve the model directly according to $\nabla_\alpha E$. For the evaluation of the cost function, we need to determine which voxels are inside or outside a given mesh. This is trivial from an algorithmic perspective, however, it is the computational bottleneck of the optimization process as this task is performed for each iteration. We addressed this issue by integrating a simple but efficient GPU accelerated voxelization algorithm introduced by Crane et al. [5] into our framework. Once all voxels inside and outside the shape are determined, we can compute the local mean values c_i and c_e , which eventually allows us to approximate the expression in (4) as follows:

$$\int_{\mathcal{S}} (f_i - f_e) N \cdot \mathcal{S}_i ds \approx \sum_{k=1}^N [f_i - f_e]_k [N]_k \cdot [\mathcal{S}_j]_k, \tag{5}$$

where k denotes the vertex number and $[\cdot]_k$ denotes the evaluation at vertex k .

3 Experiments and Results

We performed the automated segmentation for all 22 subjects, given five folds of SSM models. The only user interaction necessary was the rigid placement of the mean shape S_μ into the midbrain region. For all 22 experiments we kept the same parameter settings: the gradient descent step size was $\tau = 0.05$, the radius for the localization sphere $B_\epsilon(x)$ in (3) was set to 15 voxels ($\hat{=} 6.75\text{mm}$), and the maximum iteration number was set to 100.

In order to demonstrate that the localization of the data term (cf. (3)) is necessary, we performed another series of experiments with a standard Chan-Vese model [3]. Further, we investigated whether preprocessing the image data with a few anisotropic diffusion steps yields better results, but we could not observe a significant improvement. In all quantitative evaluations, only slices with manual ground truth were considered, since those were the only slices in which the SN was clearly visible and manually segmentable by the medical expert, i.e. we restrict our evaluation to those slices with medical relevance to the diagnostic problem.

Regarding the midbrain segmentation, we calculated the DICE coefficient of overlap between ground truth segmentations and the segmentations obtained by our method. We can observe in Fig. 2 that using the localized data term improves the segmentation results significantly. Despite the poor image quality, the median of the DICE overlap of midbrain voxels across 22 subjects is 0.83, which means that in 50% of all cases the DICE coefficient is *at least* 0.83. In contrast to this, when the un-localized data term is used, we only achieve a median DICE of only 0.55, again supporting the assumption that global foreground and background statistics are not valid in the modality of ultrasound.

In order to evaluate the quality of the segmentation with respect to the diagnostic problem, we calculated the True Positive Rate (TPR) for SN segmentations, i.e. how many SN voxels were retained within the mesh after segmentation. We found that a median of 89% of SN voxels is retained by the midbrain ROI. We also post-processed all obtained segmentations with a dilation by one voxel ($\hat{=} 0.45\text{mm}$), because we observed that the expert segmentation was systematically similar in shape to our segmentation, but segmented slightly further outwards, i.e. more into the hyper-echogenic regions surrounding the midbrain. In contrast, our segmentation converges slightly before these hyper-echogenic regions are reached. As we perform a ROI segmentation, a dilation of one voxel does not contradict the overall purpose. However, one should note that we only use dilation as one possible selection for a ROI, while our proposed method does not rely on dilation. This post-processing raises the median to 95%, i.e. half of our segmentations are able to retain more than 95% of voxels with diagnostic relevance, and five volumes achieve a perfect preservation of SN voxels after segmentation. Moreover, a dilation also improves the automatic midbrain segmentation, yielding a median DICE value of 0.86, which again shows that our segmentation is very similar in shape to the expert opinion.

In addition to the quantitative evaluation presented in Fig. 2, we want to compare one of the poorly segmented cases (case 11) to two other cases with

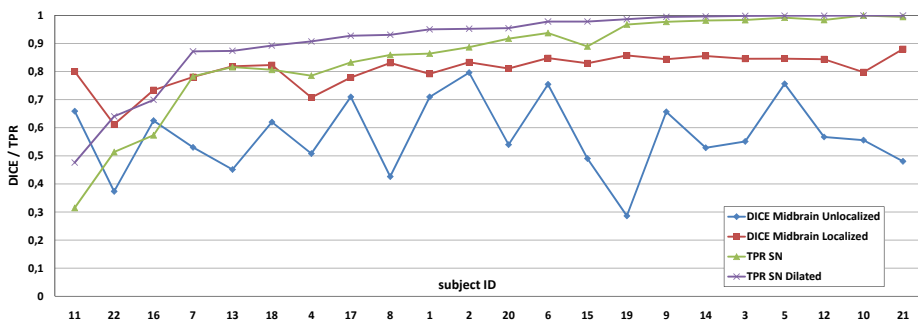


Fig. 2. DICE coefficients for overlap of our automatic segmentation with the midbrain (left) and True Positive Rate (TPR) with the substantia nigra (right). Subject indices are ordered by TPR SN Dilated, the percentage of voxels with diagnostic value retained.

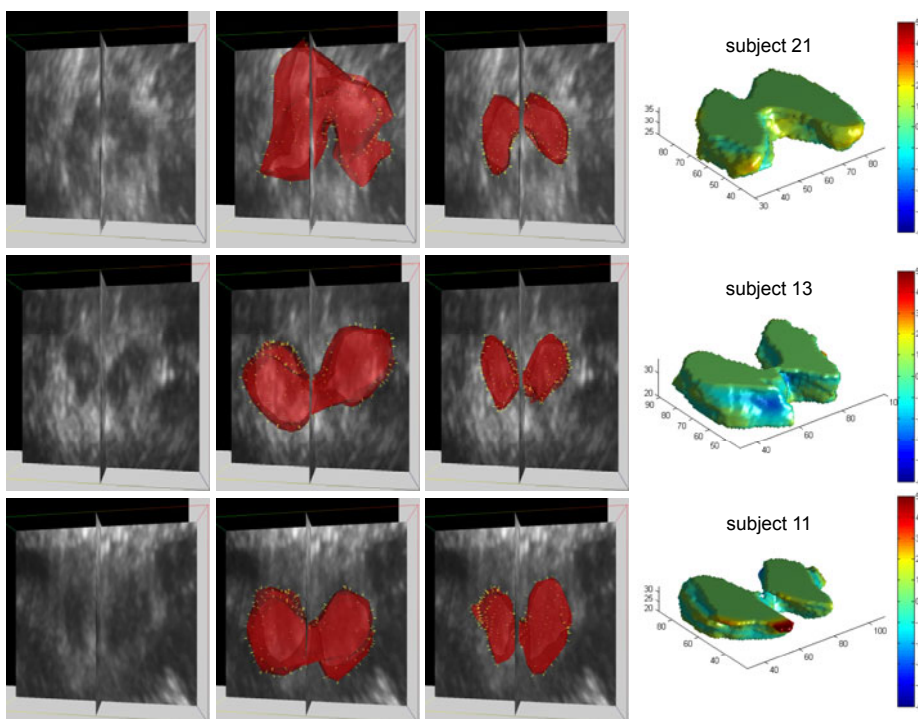


Fig. 3. Exemplary Segmentation Results: Rows: data from three subjects. Columns: sample slice through volume with midbrain visible (left), segmentation result without data term localization (middle left), segmentation result with localization (middle right), mesh surface distance map between result and ground truth (colorbar in mm).

excellent (case 21) and medium (case 13) segmentation results (cf. Fig. 2), which serve as good examples for the performance of the algorithm on the remaining data sets. The upper, middle and bottom row in Fig. 3 show cases 21, 13, and 11, respectively. The last column of the figure shows mesh surface distance maps between the ground truth segmentation and the final automatic segmentation with localized data term, showing that the automatically segmented shape corresponds well with the ground truth. The reason for the relatively poor performance in case 11 can be explained by the comparatively bad image quality and the rather unusual shape of the midbrain, making also the manual segmentation very difficult - even for a medical expert. The same holds true for cases 16 and 22.

4 Discussion and Conclusion

We have presented a robust and largely automated method for segmentation of the midbrain in 3D TC-US, which is, to the best of our knowledge, the first approach for *volumetric* segmentation of the midbrain from TC-US. The performed experiments clearly demonstrate that the segmentation performance is consistently high across 19 out of 22 subjects, although the quality of US volumes differs highly due to different thicknesses of the temporal bone windows, which proves the robustness of our method. As the image quality also depends on the used US settings and hardware, the localized region-based data term provides the advantage that it does not depend on these factors, in contrast to an Active Appearance Model for instance. Thus, the presented method can be readily applied to any form of 3D B-mode volume generation, such as wobbler probes or 2D matrix arrays.

In terms of usability, the proposed semi-automatic segmentation method reduces the overall segmentation time from approximately 20 minutes per patient for manual segmentation to around 1.5 minutes. As an accurate, robust, and user-friendly midbrain segmentation from 3D TC-US is of high importance for a subsequent segmentation *and* classification of the SN, we strongly believe that an important step has been made towards a computer aided diagnosis of Parkinson's disease, possibly improving the chance for early detection of the disease and early onset of therapy for affected patients in future.

Acknowledgments. This work was partly sponsored by EU grant FP7-ICT-2009-6-270460 ACTIVE.

References

1. Becker, G., Seufert, J., Bogdahn, U., Reichmann, H., Reiners, K.: Degeneration of substantia nigra in chronic parkinson's disease visualized by transcranial color-coded real-time sonography. *Neurology* 45(1), 182–184 (1995)
2. Berg, D., Godau, J., Walter, U.: Transcranial sonography in movement disorders. *Lancet Neurology* 7, 1044–1055 (2008)
3. Chan, T., Vese, L.: Active contours without edges. *IEEE Transactions on Image Processing* 10(2), 266–277 (2001)

4. Chen, L., Seidel, G., Mertins, A.: Multiple feature extraction for early parkinson risk assessment based on transcranial sonography image. In: 2010 17th IEEE International Conference on Image Processing (ICIP), pp. 2277–2280 (2010)
5. Crane, K., Llamas, I., Tariq, S.: Real-time simulation and rendering of 3d fluids. In: Nguyen, H. (ed.) GPU Gems 3, ch. 30. Addison Wesley Professional, Reading (2007)
6. Engel, K., Toennies, K.D.: Segmentation of the midbrain in transcranial sonographies using a two-component deformable model. *Annals of the BMVA* (4), 1–12 (2009)
7. Fang, Q., Boas, D.: Tetrahedral mesh generation from volumetric binary and grayscale images. In: IEEE International Symposium on Biomedical Imaging: From Nano to Macro, ISBI 2009, pp.1142–1145 (2009)
8. Heimann, T., Meinzer, H.P.: Statistical shape models for 3d medical image segmentation: A review. *Medical Image Analysis* 13(4), 543–563 (2009)
9. Ivancevich, N., Dahl, J., Trahey, G., Smith, S.: Phase-aberration correction with a 3-d ultrasound scanner: feasibility study. *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control* 53(8), 1432–1439 (2006)
10. Kier, C., Cyrus, C., Seidel, G., Hofmann, U.G., Aach, T.: Segmenting the substantia nigra in ultrasound images for early diagnosis of parkinson’s disease. *International Journal of Computer Assisted Radiology and Surgery* 2(S1), 83–85 (2007)
11. Lankton, S., Tannenbaum, A.: Localizing region-based active contours. *IEEE Transactions on Image Processing* 17(11), 2029–2039 (2008)
12. Michaeli, S., Oz, G., Sorce, D., Garwood, M., Ugurbil, K., Majestic, S.: Assessment of brain iron and neuronal integrity in patients with Parkinson’s disease using novel MRI contrasts. *Movement Disorders: Official Journal of the Movement Disorder Society* 22(3), 334–340 (2007)
13. Shen, L., Farid, H., McPeck, M.A.: Modeling three-dimensional morphological structures using spherical harmonics. *Evolution: International Journal of Organic Evolution* 63(4), 1003–1016 (2009)
14. Slabaugh, G., Unal, G.: Active polyhedron: surface evolution theory applied to deformable meshes. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005, vol. 2, pp. 84–91 (2005)
15. Vlaar, A., de Nijs, T., van Kroonenburgh, M., Mess, W., Winogrodzka, A., Tromp, S., Weber, W.: The predictive value of transcranial duplex sonography for the clinical diagnosis in undiagnosed parkinsonian syndromes: comparison with spect scans. *BMC Neurology* 8(1), 42 (2008)
16. Walter, U., Dressler, D., Probst, T., Wolters, A., Abu-Mugheisib, M., Wittstock, M., Benecke, R.: Transcranial brain sonography findings in discriminating between parkinsonism and idiopathic parkinson disease. *Archives of Neurology* 64(11), 1635–1640 (2007)
17. Wein, W., Pache, F., Röper, B., Navab, N.: Backward-warping ultrasound reconstruction for improving diagnostic value and registration. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4191, pp. 750–757. Springer, Heidelberg (2006)

Order Preserving and Shape Prior Constrained Intra-retinal Layer Segmentation in Optical Coherence Tomography

Fabian Rathke¹, Stefan Schmidt², and Christoph Schnörr^{1,2}

¹ University of Heidelberg, Image & Pattern Analysis Group (IPA)

² Heidelberg Collaboratory for Image Processing (HCI)

Abstract. We present a probabilistic approach to the segmentation of OCT scans of retinal tissue. By combining discrete exact inference and a global shape prior, accurate segmentations are computed that preserve the physiological order of intra-retinal layers. A major part of the computations can be performed in parallel. The evaluation reveals robustness against speckle noise, shadowing caused by blood vessels, and other scan artifacts.

1 Introduction

Over the last years Optical Coherence Tomography (OCT) has become a key technique for non-invasive diagnostic retina imaging. By measuring the backscattering of light, OCT enables to produce high-resolution 2-D and 3-D scans of retinal tissues. Quantitative measurement of the intra-retinal layers plays a central role for the early diagnosis of diseases like glaucoma or age-related macular degeneration. Since manual segmentation is tedious and time-consuming, there is a high demand for automated algorithms – see Fig. [1](#)

Related Work. The literature on segmentation is vast. Closely related work includes *heuristics* to perform 1-D edge detection and to connect candidate points across image columns into consecutive boundaries, e.g. [\[1\]](#). Another series of papers reformulate the segmentation problem as a *graph cut* problem, see [\[2\]](#) and references therein. By construction of the graph, the ordering of layers as well as the smoothness of boundaries are enforced. Edge weights are defined in terms of simple pixel-wise intensity and gradient based features. The approach takes several hours per 3-D volume. Recently, *active contour* approaches [\[3\]](#) were applied to OCT-segmentation. A functional is minimized that enforces circular boundaries (shape prior) and constant intensity within layers as well as smooth boundaries. Albeit being very fast (1s per 2-D scan), this approach requires the user to manually select the parameters.

Contribution. We present a novel probabilistic approach to the OCT segmentation problem. A global shape prior and a local appearance model are combined in a discrete graphical model which is solved for each image column separately,

thus allowing for order preserving, exact and fast parallel inference. Few updates of the model parameters by iterative conditioning suffice to incorporate the global shape prior, trained offline using ground truth data. Evaluations using independent test data yield accurate segmentations and show robustness against strong speckle noise, shadowing caused by blood vessels and other scan artifacts.

Organization. In the next section, we present an OCT image model comprising a local appearance model and a global shape prior. We describe in Sect. 3 the coupling of both models. Sect. 4 introduces the dataset we used for evaluation and reports empirical results. We conclude in Sect. 5.

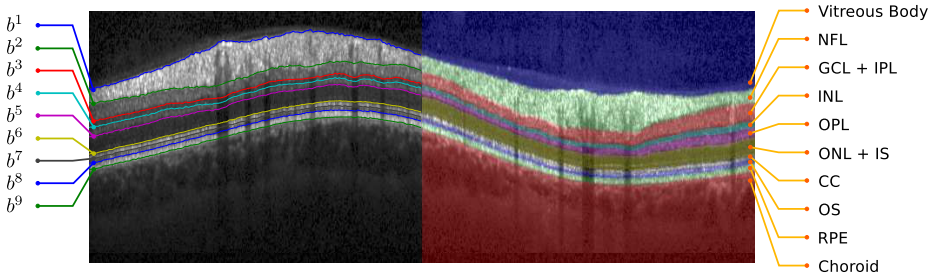


Fig. 1. An OCT B-Scan with the segmentation output of our approach: The right half shows labels l^1, \dots, l^{10} in different colors, the left half depicts the corresponding boundaries b^1, \dots, b^9 . The full names for layers l^2, \dots, l^9 are: nerve fiber layer (NFL), ganglion cell layer and inner plexiform layer (GCL + IPL), inner nuclear layer (INL), outer plexiform layer (OPL), outer nuclear layer and inner segment (ONL + IS), connecting cilia (CC), outer segment (OS), retinal pigment epithelium (RPE).

2 An OCT Image Model

We model

- in Sect. 2.1 the *local appearance* in a given $N \times M$ image I of boundaries b^1, \dots, b^9 , and corresponding layers l^2, \dots, l^9 located in between, and layers l^1, l^{10} located above b^1 and below b^9 , respectively (all shown in Fig. 1);
- in Sect. 2.2 a *global joint shape prior* for all boundaries b^1, \dots, b^9 .

2.1 Local Appearance Model

We model the appearance of pixel values I_{ij} in terms of Gaussian Markov Random Fields (GMRFs) [4] for the corresponding patches $s(i, j)$ of size 3×15 around pixel (i, j) .

For each class $k \in \{l^1, \dots, l^{10}, b^1, \dots, b^9\}$, we draw 1000 sample patches from labeled training images and convert them into vectors $s_i^k, i = 1, \dots, 1000$, of size 1×45 . Using these empirical data, we estimate for each k a class-specific density $\mathcal{N}(s; \mu_k, \Theta_k^{-1})$ with mean parameter μ_k and *sparse* precision matrix Θ_k , by applying a lasso penalty [5] as a regularizer.

We assign class variables m_{ij} to all pixels (i, j) . Given an image I , we *define* the class-conditional likelihood of I_{ij} – with slight abuse of notation – to be the likelihood of the corresponding patch

$$p(I_{ij}|m_{ij} = k) := \mathcal{N}(s(i, j)|\mu_k, \Theta_k^{-1}), \quad k \in \{l^1, \dots, l^{10}, b^1, \dots, b^9\}. \quad (1)$$

2.2 Global Shape Model

As model of the typical shape variation of layers due to both biological variability as well as to the image formation process¹, we adopt a joint Gaussian distribution of the continuous height values of all boundaries $\{b^1, \dots, b^9\}$ for all image columns j , that we denote by the $9M$ -dimensional vector $b = (b_j^n)_{n=1, \dots, 9; j=1, \dots, M}$. Hence,

$$p(b) = \mathcal{N}(b; \mu, \Sigma). \quad (2)$$

We regularize the estimation of this high-dimensional model by Probabilistic Principal Component Analysis (PPCA) [6] with a preset number k_b of eigenmodes, which yields the representation $\Sigma = WW^T + \sigma^2 I$ with a low-rank matrix W . Given the spectral decomposition of the empirical covariance matrix estimate $\tilde{\Sigma} = U\Lambda U^T$, with eigenvalues $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_{9M})$ arranged in decreasing order, the maximum likelihood estimates of the shape prior parameters σ^2, W are given by $\sigma^2 = (9M - k_b)^{-1} \sum_{i=k_b+1}^{9M} \lambda_i$ and $W = U_b(\Lambda_b - \sigma^2 I)^{1/2}$, where U_b and Λ_b denote the submatrices of U and Λ corresponding to the k_b largest eigenvalues.

3 Model Fusion and Inference

We fuse our models of appearance and shape

- in Sect. 3.1 to obtain a discrete graphical model, for which globally optimal inference (determining positions of layer boundaries) can be efficiently done in parallel for all image columns, taking the order of layers and *marginal* shape prior knowledge into account;
- by incorporating the full global shape prior knowledge *across* image columns in two alternative ways, as described in Sect. 3.2.

In a preprocessing step, we compute local class variable distributions based on the local appearance model (II),

$$p(m_{ij}|I_{ij}) = \frac{p(I_{ij}|m_{ij})p(m_{ij})}{\sum_{m_{ij}} p(I_{ij}|m_{ij})p(m_{ij})} = \frac{p(I_{ij}, m_{ij})}{p(I_{ij})}, \quad (3)$$

using a uniform prior $p(m_{ij})$.

¹ For circular scans, a wave-like distortion pattern is observed due to the conic scanning geometry and the spherical shape of the retina, which we capture statistically rather than modelling it explicitly.

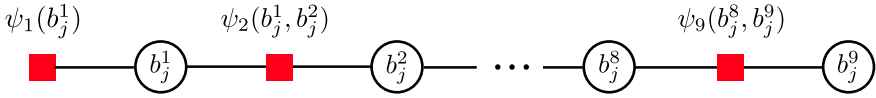


Fig. 2. The factor graph corresponding to the undirected graphical model in (4)

3.1 Order Preserving Inference and Marginal Shape Priors

For each image column $j \in \{1, \dots, M\}$, we separately set up graphical models of the form

$$p(b_j | I_j) = \frac{1}{Z} \psi_1(b_j^1) \prod_{n=2}^9 \psi_n(b_j^n, b_j^{n-1}), \tag{4}$$

in order to infer row positions of layer boundaries $b_j^n \in \{1, \dots, N\}$, for all $n = 1, \dots, 9$, conditioned on given column intensities $I_j := (I_{1j}, \dots, I_{Nj})^T$. Here, we adopt common Markovian conditional independency assumptions, leading to the factorization (4) that is graphically depicted in Fig. 2. As a consequence, we can infer in parallel for all columns, and very efficiently, layer boundary positions b_j^n .

The factors in (4) are given by

$$\psi_1(b_j^1 = i) = p(m_{ij} = b^1 | I_{ij}) p(b_j^1 = i) \tag{5a}$$

$$\psi_n(b_j^n = i, b_j^{n-1} = l) = p(m_{ij} = b^n | I_{ij}) p(b_j^n = i | b_j^{n-1} = l), \tag{5b}$$

and computed using (3) and by marginalizing out all columns but j using the global shape prior (2). Notice that the latter is a trivial operation for GMRFs (4) enabling the factorization (4) and, in turn, efficient inference.

Reading out the continuous shape prior at *discrete* row positions, and ignoring probability mass assigned to shape configurations that *violate* the natural order of layers, seems somewhat crude. Yet, this performs surprisingly well as we will demonstrate below and, after all, is computationally very efficient.

The output of the parallel inference process are expected layer positions

$$\hat{b}_j^n = \mathbb{E}[b_j^n], \quad n = 1, \dots, 9, \quad j = 1, \dots, M, \tag{6}$$

computed using the distributions (4) and the *sum-product* algorithm for marginalization.

3.2 Iterative Conditioning and Global Shape Prior

An obvious shortcoming of the parallel approach (6) is that the global shape prior (2) is only exploited by marginal distributions of b_j , separately for all columns j , in (5).

In order to overcome this drawback, we investigated two different approaches.

1. The first one simply projects \hat{b} given by (6) onto the latent PPCA subspace underlying the global shape prior (2) – see (6) for details – to obtain our final estimate \hat{b}_{proj} .

- Alternatively, the second approach iteratively updates the observation probabilities $p(m_{ij}|I_{ij})$ in (5) using the evidence (6). Specifically, using the global shape prior (2), we compute marginals for all columns j by conditioning on (6) for all other columns $-j$, and update the observation probabilities by adding weighted pseudo-observations

$$p(m_{ij} = b^n | I_{ij}) \leftarrow p(m_{ij} = b^n | I_{ij}) + \frac{p_{\text{best}} - p_{\text{pred}}}{p_{\text{best}}} \delta_{i, \tilde{b}_j^n}, \quad (7)$$

$$n = 1, \dots, 9, \quad i = 1, \dots, N$$

where $\delta_{i, \tilde{b}_j^n}$ is the Kronecker delta, and p_{pred} and p_{best} are conditional marginals evaluated for the *prediction* \hat{b}_j^n and the (in terms of the shape model) *best* boundary $\tilde{b}_j^n = \operatorname{argmax}_{i \in \{1, \dots, N\}} p(b_j^n = i | \hat{b}_{-j})$ respectively.

Subsequently, we again infer (6) based on the terms (5) updated by (7), and iterate this process until \hat{b} converges. Experiments show that this happens after few steps, and thus does not compromise computational efficiency.

4 Evaluation and Discussion

This section reports the performance of the approaches described in Sect. 2 and 3. To this end, we evaluated the following models:

- Uniform Prior*: $p(b_j^n = i | b_j^{n-1} = l) = \text{const.}, \forall i > l$, used in order to inspect how performance degrades *without* shape prior information.
- Shape Prior (SP)*: The graphical model described in Sect. 3.1
- SP + Projection*: Post-processing \hat{b} by projection onto the PPCA subspace, described in Sect. 3.2, alternative 1.
- SP + Iteration*: Post-processing \hat{b} by iterative conditioning, described in Sect. 3.2, alternative 2.

Data Aquisition. Our dataset comprises 80 circular B-scans from 80 healthy persons with given ground truth labeling. The dataset was acquired by using a Heidelberg Engineering Spectralis HRA+OCT device. Each scan has a diameter of 3.4mm and consists of $M = 768$ A-scans of depth resolution $3.87 \mu\text{m}/\text{pixel}$ ($N = 496$ pixels).

Evaluation. For each boundary we compute the unsigned, signed and squared distances between the estimate \hat{b}^n and the manual segmentation \tilde{b}_j^n (ground truth),

$$E_{\text{unsgn}}^n = \sum_{j=1}^M \hat{b}_j^n - \tilde{b}_j^n, \quad E_{\text{sgn}}^n = \sum_{j=1}^M |\hat{b}_j^n - \tilde{b}_j^n|, \quad E_{\text{sq}}^n = \sum_{j=1}^M (\hat{b}_j^n - \tilde{b}_j^n)^2, \quad (8)$$

where M indicates the number of image columns (A-scans). The errors of the whole segmentation are computed as

$$E_* = (9M)^{-1} \sum_{n=1}^9 E_*^n, \quad * \in \{\text{unsgn}, \text{sgn}, \text{sq}\}. \quad (9)$$

A 10-fold *nested* cross-validation over the lasso parameter (Sec. 2.1) was performed: First, the dataset was split into 10 subsets. For each subset (the *hold-out* set), the optimal parameter was determined by an *inner* 5-fold cross-validation on the remaining 9 subsets. Given this parameter, a model was trained using all 9 subsets and performance was evaluated on the hold-out set. Fig. 3 reports results in terms of E_{unsgn} , E_{sgn} and E_{sq} with error bars indicating standard deviation. Fig. 4 shows characteristic segmentation results for all four approaches in comparison.

Segmentation with the uniform prior performs worst, indicating the necessity of using a shape prior. Adding prior statistical information about the relative distances of layer boundaries (model 2. listed above) within each column boosts performance significantly. For many columns this approach finds the true boundary positions. Nevertheless, for columns with specific appearance, e.g. caused by blood vessels, segmentation may fail (see Fig. 4b). This reveals the lack of communication across image columns. Adding either PPCA projection (model 3.) or iterative conditioning (model 4.) as a post-processing step can resolve such issues (Fig. 4c and d).

As an illustration, Fig. 5 depicts the addition of pseudo-observations (7) for boundaries 6 and 9 in column j for the first round of iterative conditioning. Depending on the distance between \tilde{b}_j^n (marked as green points) and \hat{b}_j^n , strong (boundary 6) or weak (boundary 9) pseudo-observations are added. As a consequence, estimates \hat{b}_j^n that rely on weak observations and/or differ significantly from \tilde{b}_j^n are corrected by the shape prior. While the iterative approach is working well for the lower boundaries $b^6 - b^9$, it may happen that estimates for $b^2 - b^5$ are too smooth, as shown for b^2 in Fig. 4d, where the shape prior somewhat tends to overrule the appearance model.

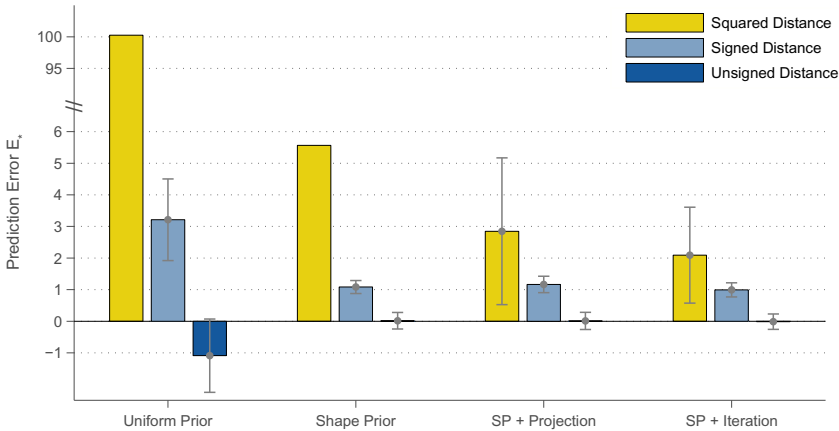


Fig. 3. Error means and standard deviation for all approaches 1., . . . , 4., evaluated for all 80 scans by means of a 10-fold nested cross-validation. The standard deviations of the squared error for the uniform prior and the shape prior model are 128.34 and 8.75 respectively.

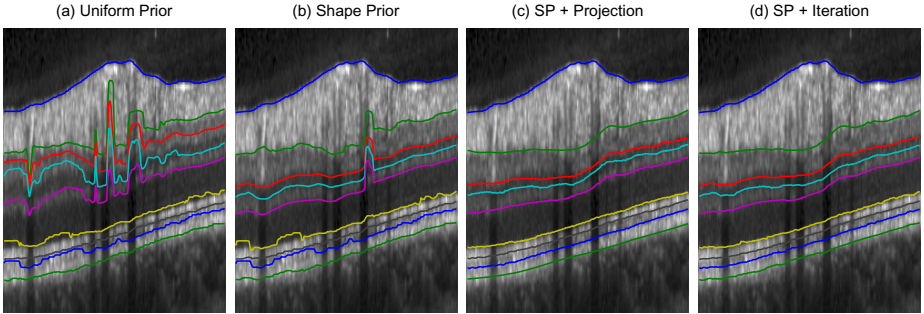


Fig. 4. Close-up view of segmentation results for the approaches 1., . . . , 4.. Left panel: no shape prior. Adding the shape prior significantly improves segmentation performance, especially for boundaries $b^2 - b^5$, but still may fail locally (2nd panel from left). Iteratively enforcing the full shape prior leads to good results (3rd and 4th panel).

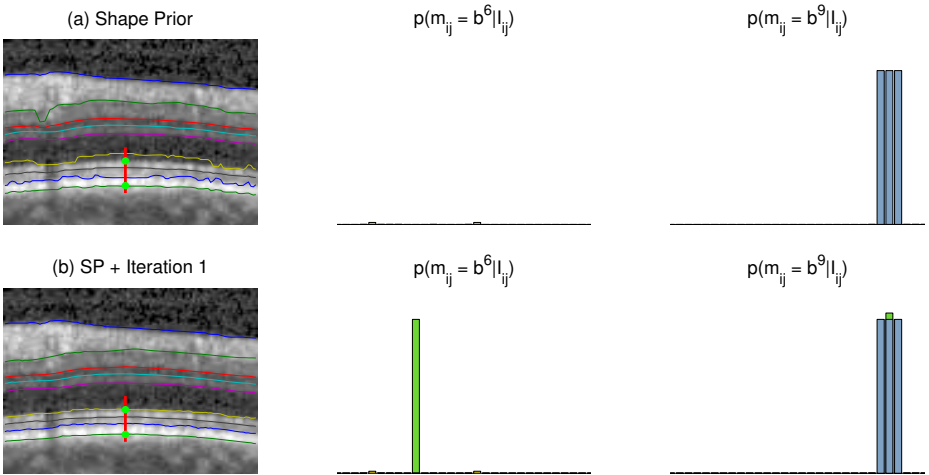


Fig. 5. The left panels show segmentations before (a) and after (b) the first step of iterative conditioning. Green points indicate \tilde{b}_j^n for $n = 6, 9$ (see Sect. 3.2). The middle and right panels show the corresponding observation probabilities of both boundaries for the red marked segment of column j . Green bars indicate the magnitude of pseudo-observations added in (7), being stronger if the estimate \hat{b}_j^n differs significantly from \tilde{b}_j^n .

Our current implementation in Matlab requires ~ 15 seconds per 2D-scan on an Intel Core 2 Quad Q9550 with 2.83 Ghz.

5 Conclusion and Further Work

This work presented a novel probabilistic model for the segmentation of circular OCT scans. It combines a local appearance model and a global shape model using

a graphical model that enables very efficient computational inference. Comparison of different model variants has shown that utilizing the full shape prior performs best. The approach works without user interaction, is robust against appearance artifacts, and returns accurate segmentation results.

Our future work will investigate computationally tractable extensions that enable horizontal coupling of image columns not only by the appearance model but also by the shape prior. Furthermore, we will apply our approach to other types of OCT scans, e.g. 3D volumes or scans from the fovea region. The extension to 3D is straightforward: The shape prior as well as the texture model can be trained on 3D patches/boundaries, while the graphical model in Sec. 3 can be augmented to include columns (A-scans) of all slices in the volume. Finally, we will focus on discriminating non-/pathological scans and the development of statistical tools for assessing such images locally.

Acknowledgments. The authors would like to thank Heidelberg Engineering, for providing the OCT dataset as well as their employees Karsten Müller, Dr. Thomas Fendrich and Dr. Tilman Otto for helpful discussions.

This work has been supported by the German Research Foundation (DFG) within the programme “Spatio-/Temporal Graphical Models and Applications in Image Analysis”, grant GRK 1653.

References

1. Koozekanani, D., Boyer, K., Roberts, C.: Retinal thickness measurements from optical coherence tomography using a Markov boundary model. *IEEE T. Med. Imaging* 20(9), 900–916 (2001)
2. Garvin, M., Abramoff, M., Wu, X., Russell, S., Burns, T., Sonka, M.: Automated 3-D intraretinal layer segmentation of macular spectral-domain optical coherence tomography images. *IEEE T. Med. Imaging* 28(9), 1436–1447 (2009)
3. Yazdanpanah, A., Hamarneh, G., Smith, B.R., Sarunic, M.V.: Segmentation of intra-retinal layers from optical coherence tomography images using an active contour approach. *IEEE T. Med. Imaging* 30(2), 484–496 (2011)
4. Rue, H., Held, L.: *Gaussian Markov Random Fields: Theory and Applications*. Chapman & Hall/CRC, London (2005)
5. Friedman, J., Hastie, T., Tibshirani, R.: Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* 9(3), 432–441 (2008)
6. Tipping, M.E., Bishop, C.M.: Probabilistic principal component analysis. *J. R. Stat. Soc.* 61(3), 611–622 (1999)

Adaptive Neonate Brain Segmentation

M. Jorge Cardoso¹, Andrew Melbourne¹, Giles S. Kendall²,
Marc Modat¹, Cornelia F. Hagmann², Nicola J. Robertson²,
Neil Marlow², and Sebastien Ourselin¹

¹ Centre for Medical Image Computing (CMIC), University College London, UK

² Academic Neonatology, EGA UCL Institute for Women's Health, London, UK

Abstract. Babies born prematurely are at increased risk of adverse neurodevelopmental outcomes. Recent advances suggest that measurement of brain volumes can help in defining biomarkers for neurodevelopmental outcome. These techniques rely on an accurate segmentation of the MRI data. However, due to lack of contrast, partial volume (PV) effect, the existence of both hypo- and hyper-intensities and significant natural and pathological anatomical variability, the segmentation of neonatal brain MRI is challenging. We propose a pipeline for image segmentation that uses a novel multi-model Maximum a posteriori Expectation Maximisation (MAP-EM) segmentation algorithm with a prior over both intensities and the tissue proportions, a B0 inhomogeneity correction, and a spatial homogeneity term through the use of a Markov Random Field. This robust and adaptive technique enables the segmentation of images with high anatomical disparity from a normal population. Furthermore, the proposed method implicitly models Partial Volume, mitigating the problem of neonatal white/grey matter intensity inversion. Experiments performed on a clinical cohort show expected statistically significant correlations with gestational age at birth and birthweight. Furthermore, the proposed method obtains statistically significant improvements in Dice scores when compared to the a Maximum Likelihood EM algorithm.

1 Introduction

Preterm birth (below 32 weeks gestation) is associated with significant cognitive and neuromotor impairments, the frequency and severity of which increase with decreasing gestational age [1]. Much research has been focused on the development of biomarkers of neurological injury in an attempt to identify babies at increased risk of adverse neurodevelopmental outcomes. Recently an approach combining magnetic resonance imaging and volumetric analysis has demonstrated a correlation with neurodevelopmental outcome at 2 years [2]. Fundamental to performing volumetric and morphometric studies is the ability to classify different brain tissues.

In contrast with adults, neonatal brain MRI is complex to segment due to lack of contrast, partial volume (PV) effect, the existence of hypo- and hyper-intensities and substantial natural and pathological anatomical variability. Dynamic natural developmental variability arises due to the receding germinal

matrix and proceeding myelination from posterior to anterior. Cortical complexity also increases as the cerebral grey matter surface becomes more convoluted. Signal intensities of grey and white matter (GM/WM) on neonatal T1 and T2 MRI are characteristically reversed and contrast between the two tissue types is progressively lost during the first few months of life as a result of the myelination process resulting in low contrast-to-noise (CNR) and signal-to-noise ratio (SNR) images. Recent developments in neonatal brain segmentation have used template atlases [3] to mitigate the low SNR and CNR of the images. These methods rely on the registration of templates in order to propagate knowledge of a particular population to a patient specific space. This is problematic in pathological cases, as their anatomy may be markedly different from the one derived from a normal population atlas (see figure 1). Zhuang et al. [4] proposed a simpler intensity based classification method based on K-means and augmented by a population atlas and Yu et al. [5] proposed a Parzen windows based Hidden Markov Random Field algorithm, optimised by an Expectation-Maximisation algorithm, again augmented by a population atlas. Both these methods will have problems in severely diseased patients as the propagation of the templates by image registration may not be able to cope with the variability. Furthermore, the contrast inversion causes a layer of PV corrupted voxels on the CSF/GM boundary to have similar intensities to pure WM. Xue et al. [6] tries to ameliorate this problem through morphological operations without implicitly modelling PV.

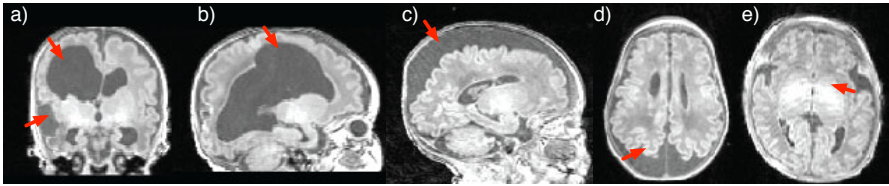


Fig. 1. Red arrows pointing to the problematic areas. Pathological areas that differ from the normal population in patients with a) and b) ventriculomegaly, c) and d) excessive sub-arachnoid CSF. e) Movement artefacts and hyper-intensities.

In this paper, we propose a new segmentation pipeline incorporating a novel *Maximum a Posteriori* Expectation-Maximization (MAP-EM) based probabilistic segmentation technique that includes intensity non-uniformity (INU) correction, spatial dependence via a Markov Random Field (MRF) and PV containing voxels. The pipeline also iteratively relaxes normal population priors, thus enabling their adaptation to pathological cases. The segmentation results are then used to iteratively update the registration from the template space to the subject, improving the prior alignment, the segmentation and the skull-stripping.

2 Methods

2.1 MAP Expectation-Maximization Segmentation

Assuming either single modality or coregistered multimodal datasets, let $i \in \{1, 2, \dots, n\}$ index the n voxels of an image domain, with its intensities forming

a feature vector $y_i \in \mathcal{R}^m$. For K tissue types, let z_i denote the tissue type of voxel i . Thus, $z_i = e_k$ for some k , $1 \leq k \leq K$ where e_k is a unit vector with the k th component equal to one and all the other components equal to zero. Let each tissue type be described as having multivariate normally distributed intensities with mean vector $\mu_k \in \mathcal{R}^m$ and respective covariance matrix $\Lambda_k \in \mathcal{R}^{m \times m}$ grouped in $\theta_k = \{\mu_k, \Lambda_k\}$, corrupted by a multiplicative bias field. In this, work, the bias field is modelled as a linear combination $c' \phi$ of J smoothly varying basis functions $\phi(x) = \{\phi(x)_1, \dots, \phi(x)_J\}$, where x denotes the spatial position and $c = \{c_1, \dots, c_m\}'$ with c_m as a vector of coefficients for channel m . For mathematical convenience and similarly to [7] and [8], we assume that the intensity of the voxels that belong to class k are normally distributed after log transformation. This log transformation of the data makes the multiplicative bias field additive, ameliorating problems with numerical stability and enabling the existence of a linear least square solution for the coefficient optimisation.

Let $\Phi_y = \{\theta_1, \theta_2, \dots, \theta_K, C\}$ represent the overall model parameters. Assuming an *a priori* distribution over these parameters, the problem can be formalised as a *Maximum a Posteriori* (MAP) estimation of Φ_y by

$$\hat{\Phi}_y = \arg \max_{\Phi} f(y | \Phi_y) g(\Phi_y) \tag{1}$$

with $f(y | \Phi_y) = \prod_i \prod_k f(y_i | z_i = e_k, \Phi_y) f(z_i = e_k | p_{\mathcal{N}_i} \Phi_z)$ and $g(\Phi_y)$ as a semi-conjugate gaussian prior over the mean. These priors are used to constrain the space of solutions of the mean, increasing the algorithm’s robustness in complex segmentation cases. Here, the parameter μ_k , defined as

$$g(\mu_k) = \left((2\pi)^{\frac{m}{2}} | \Lambda_k^* |^{\frac{1}{2}} \right)^{-1} e^{-\frac{1}{2} (\mu_k - \mu_k^*)' (\Lambda_k^*)^{-1} (\mu_k - \mu_k^*)} \tag{2}$$

with $| \Lambda_k^* |$ as the determinant of Λ_k^* and μ_k^* and Λ_k^* as priors over the parameter μ_k . The need for this prior over the means will become apparent in Section 2.2.

Here, $f(y | \Phi_y)$ can be seen as a mixture of multivariate normal distributions, thus $f(y_i | z_i = e_k, \Phi_y) = G_{\Lambda_k}(y_i - \mu_k - c' \phi(x_i))$, where $G_{\Lambda_k}(x)$ denotes a multivariate zero-mean normal distribution with standard deviation Λ_k , defined as $G_{\Lambda_k}(x) = ((2\pi)^{\frac{m}{2}} | \Lambda_k |^{\frac{1}{2}})^{-1} e^{-\frac{1}{2} x' (\Lambda_k)^{-1} x}$. Thus, the *Maximum a Posteriori* (MAP) parameters for Φ_y can be found by maximisation of Eq. 1 giving the following update equations for the model parameters:

$$\begin{aligned} \mu_k^{(t+1)} &= \left(\mu_k^* \Lambda_k^{*-1} + (\sum_{i=1}^n p_{ik} \bar{y}_{ik}) \Lambda_k^{(t)-1} \right) \left(\Lambda_k^{*-1} + (\sum_{i=1}^n p_{ik}) \Lambda_k^{(t)-1} \right)^{-1} \\ \Lambda_k^{(t+1)} &= \sum_{i=1}^n p_{ik} \left(\bar{y}_{ik} - \mu_k^{(t+1)} \right)' \left(\bar{y}_{ik} - \mu_k^{(t+1)} \right) \left(\sum_{i=1}^n p_{ik} \right)^{-1} \end{aligned}$$

where $\bar{y}_{ik} = y_i - (c^{(t)})' \phi(x_i)$ and

$$p_{ik}^{(t+1)} = f(y_i | z_i = e_k, \Phi_y^{(t)}) f(z_i = e_k) \left(\sum_{j=1}^K f(y_i | z_i = e_j, \Phi_y^{(t)}) f(z_i = e_j) \right)^{-1}$$

is the responsibility at the index i and class k and t denotes the iteration number. The estimation of $c^{(t+1)}$ under a multimodel scheme is provided in [9].

Spatial Regularisation: Due to the low signal and contrast to noise ratio of the images under analysis, the assumption of statistical independence between neighbouring voxels makes the segmentation prone to noise and image artefacts. A spatial smoothness term by means of a Markov Random Field (MRF) is thus included. This MRF is a non-binary extension of a multiclass Potts model, and assumes that the probability that voxel i belongs to tissue k depends on its first-order neighbours \mathcal{N}_i .

Using the mean field approximation as described in [10] and [9], all previous equations still hold by setting $f(z_i = e_j) = f(z_i = e_k | p_{\mathcal{N}_i} \Phi_z, \pi_{ik})$ in Eq 2.1, where

$$f(z_i = e_k | p_{\mathcal{N}_i}, \Phi_z, \pi_{ik}) = \left(\pi_{ik} e^{-U_{\text{MRF}}(e_k | p_{\mathcal{N}_i}, \Phi_z)} \right) \left(\sum_{j=1}^K \pi_{ij} e^{-U_{\text{MRF}}(e_j | p_{\mathcal{N}_i}, \Phi_z)} \right)^{-1}$$

Here, $U_{\text{MRF}}(z_i | p_{\mathcal{N}_i}, \Phi_z)$ is an energy function dependent on the parameters $\Phi_z = \{G_{kj}, \beta\}$, where β is the overall MRF strength. Under anisotropic voxel size, the interaction between neighbours is dependant on a connection strength factor s , introduced as $s = \{s_x, s_y, s_z\} = \{\frac{1}{d_x}, \frac{1}{d_y}, \frac{1}{d_z}\}$, where d is the real-world distance between the centre of neighbouring voxels in each direction. Under this framework,

$$U_{\text{MRF}}(e_k | p_{\mathcal{N}_i}, \Phi_z) = \beta \sum_{j=1}^K G_{kj} \left(\sum_{l \in \mathcal{N}_i^x} s_x p_{lj} + \sum_{l \in \mathcal{N}_i^y} s_y p_{lj} + \sum_{l \in \mathcal{N}_i^z} s_z p_{lj} \right)$$

Relaxation of the Anatomical Priors: Due to anatomical variability and similarly to [11], we assume that the proportions π_{ik} are not known *a priori*. Instead, we consider π_{ik} as a sample drawn from a distribution derived from the statistical atlas, i.e., we consider them as a posterior of a Dirichlet distribution. Here, π_{ik} is updated at each iteration by

$$\pi_{ik} = (1 - \alpha)\omega_{ik} + \alpha(\mathcal{G}_{\sigma_{Der}} * p_{ik}) \quad (3)$$

with $\mathcal{G}_{\sigma_{Der}}$ as a gaussian kernel with standard deviation σ_{Der} and $*$ as the convolution operation, weighted by α . This updating scheme iteratively spatially relaxes π_{ik} , enabling the segmentation of pathological cases. The amount of relaxation is dependent on the parameter α controlling the mixing proportions, and σ_{Der} controlling the amount of regularisation over the posterior. For $\alpha = 0$, only the anatomical priors are used. These anatomical priors ω that incorporate probabilistic information derived from a digital brain atlas are added to the model in order to condition the posterior probabilities and indirectly also condition the model parameters. These atlases are brought into correspondence using an affine registration [12] followed by a free-form non-rigid registration algorithm [13].

Assuming skull stripped images, we model the problem with $K = 6$ classes, each one with a corresponding digital atlas prior probability for white matter

(WM), cortical grey matter (cGM), deep grey matter (dGM), cerebrospinal fluid (CSF), pons (Pon) and cerebellum (Cer) respectively at every voxel position. The anatomical priors provided by [14] are used here. The time point in the 4D atlas is chosen depending on the gestational age of the neonate under analysis.

2.2 Segmentation Pipeline

The proposed pipeline can be described as a sequence of registrations and segmentations, that acts to improve the atlas alignment, the propagation of the brain mask and consequently the final segmentation. First, the population template is registered to the patient space and both the mask and the anatomical priors are propagated using the same deformation field. Due to the low signal to noise ratio and architectural variability of the images (shown in Figure 1) the registration is not able to capture the correct deformation field on the first pass. However, at this stage, the registration is able to align the more anatomically consistent structures (skull, scalp) despite performing quite badly in situations with excessive sub-arachnoid CSF or with marked ventricular expansion.

After the first registration step, the image is segmented with the above described MAP-EM segmentation algorithm with very broad parameters for the prior relaxation. At this stage, we use $\sigma_{Der} = 6$ and $\alpha = 0.5$, enabling a great adaptation of the priors. Because the prior relaxation is an iterative process, in patients with marked anatomical disparity the tissue classes might not converge to the desired solution at this stage. For example, in Fig 2(left), the WM class converges to a gaussian with large standard deviation including both CSF and WM and the INU correction tries to compensate the intensity difference. In order to ameliorate this problem, the images are intensity normalised using a robust max-min (assuming 2 % outliers) after masking. The expected distribution of the parameter μ for each class was then obtained by manually sampling several 3x3x3 voxel cubic patches over a set of intensity normalised images. From this sampling process, the mean and standard deviation of the mean intensity over each patch was calculated and assigned to the prior μ_k^* and Λ_k^* . On multimodal images, the process is done separately for each modality as Λ_k^* is assumed to be diagonal. The inclusion of this extra term adds some *a priori* knowledge about the class specific distribution of the parameter μ

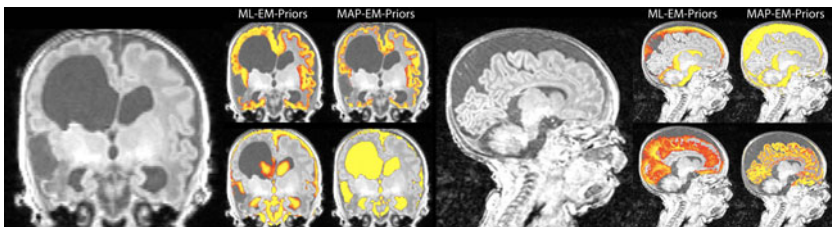


Fig. 2. Patients with either marked ventricular expansion (left) or excessive sub-arachnoid CSF (right) and their corresponding cortical GM and CSF segmentations, with and without the prior over the mean intensities.

increasing the robustness of its estimation. After the segmentation, the registration is updated by matching the brain prior (sum of all the brain structures excluding CSF) with the current brain segmentation. This allows a much less noisy and simplified registration process. Both the brain mask and the anatomical priors are re-propagated using the new deformation field. The images are segmented again using the new priors and mask, and the values of σ_{Der} and α are reduced to 2 and 0.3 respectively, resulting in a more prior driven and less adaptive segmentation.

Explicit PV modelling: At this stage, the segmentation is generally good and the priors are now correctly aligned. However, there is still a layer of WM classified voxels on the GM/CSF, Cerebellum/CSF and Pons/CSF interfaces (see Fig. 3). It is common to assume that if two tissues mix in a voxel, all mixing proportions are equally likely. The PV probability can be seen as a number of mixed Gaussians in between the two pure classes, corresponding to all the possible tissue proportions within a voxel [15]. Ruan *et al.* [16] showed that, for brain imaging and for the signal-to-noise ratio levels of current MRI systems, the density of all these PV gaussian classes can be approximated by a single gaussian with a small risk ($\alpha < 1$ for D’Agostino-Pearson normality test). Under this assumption, and similarly to [17], we use the previously estimated values of p_{ik} , μ_k and σ_k to initialise the mixed classes.

Under the assumption of Gaussian distributed classes on log-transformed data, the initial PV class Gaussian parameters can be approximated by a mixel distribution [18], with mean equal to the arithmetic weighted average of its composing class parameters weighted by the determinant of the covariance matrix of each class. Thus, $\mu_{j/k}^* = |\Lambda_j| \mu_j + |\Lambda_k| \mu_k$. Due to the multiplicative nature of the probabilities, the mixed class prior is generated as the normalised geometric mean of its composing tissue distributions p_{ij} and p_{ik} .

$$\omega_{i(j/k)}^* = \sqrt{p_{ij} p_{ik}} \quad (4)$$

and

$$\omega_{i(WM)}^* = \omega_{i(WM)} \left(1 - \sqrt{\omega_{i(GM)} \omega_{i(CSF)}}\right) \quad (5)$$

This transformation will reduce the *a priori* probability for PV containing voxels to belong to WM. The priors are then normalised in order to sum to one at each voxel position. Even though these new priors cannot be considered as priors in a strict sense as they are derived from patient specific data, they behave as such in the model. After these patient specific PV priors are created, the image is segmented again using the above described MAP-EM algorithm with the 3 extra PV containing classes (GM/CSF, Cerebellum/CSF and Pons/CSF interfaces), resulting in a much reduced amount of PV containing voxels classified as WM. Please refer to Fig. 3 for a visual assessment of the segmentation steps and their improvements.

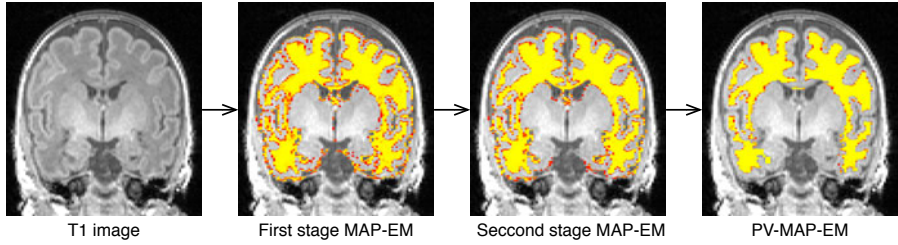


Fig. 3. From left to right: Segmentation pipeline showing the original image, the first segmentation step, the second segmentation step with the improved prior alignment and finally the PV corrected segmentation result .

3 Validation

Data: The data were acquired on a 1.5T Siemens's Avanto using $TR = 17ms$, $TE = 6ms$ and flip angle of 21° . In total 43 T1-weighted volumes are analysed with resolution $0.39 \times 0.39 \times 1mm$. The mean gestational age is 27.1 ± 2.7 weeks (range 23.1 – 32.3), mean birthweight $970 \pm 373g$ (range 540 – 2470g) and mean age at scan 40.4 ± 1.74 (range 35.7 – 44.3). The male to female ratio is 37/36.

Clinical correlation: For each segmentation we analyse correlations with GA at birth and birthweight. We additionally correlate the grey/white matter ratio. The correlation coefficient required to achieve a p-value ≤ 0.01 for this sample size (43) is $r = 0.36$. With the exception of the CSF volume, each of the segmented components correlates significantly ($r > 0.5$) with both gestational age at birth and birthweight. The absence of correlation with CSF is expected due to significant natural and pathological variation in ventricle size. There are no correlations with gender.

Quantitative and qualitative analysis: In order to assess the quality of the segmentation, the proposed segmentation was compared to manually segmented ventricles. Two subsets were chosen from the full database: one containing 15 neonates with normal anatomy and another containing 4 patients with noticeable anatomical differences (2 cases with ventriculomegaly and 2 with excessive sub-arachnoid CSF). For each dataset, a Dice score was calculated between the proposed method and the manual segmentation. The proposed algorithm was also compared to the Maximum Likelihood Expectation Maximisation (ML-EM) algorithm [9]. On the subset containing the 15 normal subjects, the the proposed method obtains a dice score of 0.908 ± 0.034 when compared to 0.725 ± 0.217 for the ML-EM, representing a statistically significant improvement ($p < 0.01$) in the accuracy of the segmentation. On the second subset, the dice scores for the proposed method are 0.92, 0.86, 0.93 and 0.94, and respectively 0.22, 0.15, 0.35, and 0.55 for the ML-EM method, showing a marked improvement in the accuracy of the segmentation in pathological cases.

4 Conclusions

We have presented a segmentation algorithm tailored specifically for neonatal T1 segmentation, using a MAP-EM algorithm with a new prior relaxation strategy combined with a semi-conjugate prior over the intensities and an implicit PV model in order to mitigate the problematic misclassified layer of PV voxels. Statistically significant correlations are shown between structural volumes and both weight at birth and gestational age. Furthermore, experiments performed on a clinical cohort also show significant improvements in segmentation accuracy, mainly in pathological cases, when compared to a ML-EM algorithm.

References

1. Marlow, N., Wolke, D., Bracewell, M.A., Samara, M.: Neurologic and developmental disability at six years of age after extremely preterm birth. *New England Journal of Medicine* 352(1), 9–19 (2005)
2. Boardman, J., Craven, C., Valappil, S., Counsell, S., Dyet, L., Rueckert, D., Aljabar, P., Rutherford, M., Chew, A., Allsop, J., Cowan, F., Edwards, A.: A common neonatal image phenotype predicts adverse neurodevelopmental outcome in children born preterm. *NeuroImage* 52(2), 409–414 (2010)
3. Weisenfeld, N.I., Warfield, S.K.: Automatic segmentation of newborn brain mri. *NeuroImage* 47(2), 564–572 (2009)
4. Song, Z., Awate, S., Licht, D., Gee, J.: Clinical neonatal brain mri segmentation using adaptive nonparametric data models. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part I. LNCS, vol. 4791, pp. 883–890. Springer, Heidelberg (2007)
5. Yu, X., Zhang, Y., Lasky, R.E., Parikh, N.A., Narayana, P.A.: Comprehensive brain mri segmentation in high risk preterm newborns. *PLoS ONE* 5(11) (2010)
6. Xue, H., Srinivasan, L., Jiang, S., Rutherford, M., Edwards, A.D., Rueckert, D., Hahn, J.V.: Automatic segmentation and reconstruction of the cortex from neonatal mri. *NeuroImage* 38(3), 461–477 (2007)
7. Wells, W., Grimson, W.E., Kikinis, R., Jolesz, F.A.: Adaptive segmentation of MRI data. *IEEE Transactions on Medical Imaging* 15(4), 429–442 (1996)
8. Zhang, Y., Brady, M., Smith, S.M.: Segmentation of brain MR images through a hidden markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging* 20(1), 45–57 (2001)
9. Van Leemput, K., Maes, F., Vandermeulen, D., Suetens, P.: Automated model-based tissue classification of MR images of the brain. *IEEE TMI* 18(10) (1999)
10. Zhang, J.: The mean field theory in em procedures for markov random fields. *IEEE Transactions on Signal Processing* 40(10), 2570–2583 (1992)
11. Shiee, N., Bazin, P.L., Cuzzocreo, J.L., Blitz, A., Pham, D.L.: Segmentation of brain images using adaptive atlases with application to ventriculomegaly
12. Ourselin, S., Roche, A., Prima, S., Ayache, N.: Block matching: A general framework to improve robustness of rigid registration of medical images. In: Delp, S.L., DiGoia, A.M., Jaramaz, B. (eds.) MICCAI 2000. LNCS, vol. 1935, pp. 557–566. Springer, Heidelberg (2000)
13. Modat, M., Ridgway, G., Taylor, Z., Lehmann, M., Barnes, J., Hawkes, D., Fox, N., Ourselin, S.: Fast free-form deformation using graphics processing units. *Computer Methods and Programs in Biomedicine* (October 2009)

14. Kuklisova-Murgasova, M., Aljabar, P., Srinivasan, L., Counsell, S.J., Gousias, I.S., Boardman, J.P., Rutherford, M.A., Edwards, A.D., Hajnal, J.V., Rueckert, D.: A dynamic 4d probabilistic atlas of the developing brain. *NeuroImage* 54(4) (2011)
15. Van Leemput, K., Maes, F., Vandermeulen, D., Suetens, P.: A unifying framework for partial volume segmentation of brain MR images. *IEEE Transactions on Medical Imaging* 22(1), 105–119 (2003)
16. Ruan, S., Jaggi, C., Fadili, J., Bloyet, D.: Brain tissue classification of magnetic resonance images using partial volume modeling. *IEEE TMI* 19(12) (December 2000)
17. Cardoso, M.J., Clarkson, M.J., Ridgway, G.R., Modat, M., Fox, N.C., Ourselin, S., The Alzheimer's Disease Neuroimaging Initiative: LoAd: A locally adaptive cortical segmentation algorithm. *NeuroImage* 56(3), 1386–1397 (2011)
18. Kitamoto, A., Takagi, M.: Image classification using probabilistic models that reflect the internal structure of mixels. *Pattern Analysis and Applications* 2 (1999)

Renal Cortex Segmentation Using Optimal Surface Search with Novel Graph Construction

Xiuli Li¹, Xinjian Chen², Jianhua Yao², Xing Zhang¹, and Jie Tian^{1,*}

¹ Institute of Automation, Chinese Academy of Sciences, China
tian@ieee.org

² Radiology and Imaging Sciences Department,
Clinical Center, National Institute of Health, USA

Abstract. In this paper, we propose a novel approach to solve the renal cortex segmentation problem, which has rarely been studied. In this study, the renal cortex segmentation problem is handled as a multiple-surfaces extraction problem, which is solved using the optimal surface search method. We propose a novel graph construction scheme in the optimal surface search to better accommodate multiple surfaces. Different surface sub-graphs are constructed according to their properties, and inter-surface relationships are also modeled in the graph. The proposed method was tested on 17 clinical CT datasets. The true positive volume fraction (TPVF) and false positive volume fraction (FPVF) are 74.10% and 0.08%, respectively. The experimental results demonstrate the effectiveness of the proposed method.

1 Introduction

Image segmentation is a fundamental and challenging issue in the area of medical image analysis and computer-aided diagnosis. The kidney volume and renal cortex thickness are important clinical indications for renal function and are valuable for nephrologists. As shown in Fig. 1, the kidney consists of the renal cortex, renal medulla and renal pelvis. The renal cortex forms a continuous smooth outer zone with a number of projections (renal columns). Several kidney segmentation methods have been proposed in the literature [1-3]. They focused on mostly the whole kidney and seldom referred to the specific structures such as the renal column and renal cortex. Since the renal cortex thickness information plays an important role in renal function assessment, the segmentation of the renal cortex is desirable.

Special characteristics of the kidney's anatomical structure, such as (1) semi-closed structure around the renal pelvis, (2) weak boundaries between adjacent organs (e.g. liver, spine and muscles), and (3) the intensity proximity of the renal cortex and renal column (see Fig. 1), make renal cortex segmentation a challenging problem. Traditional methods such as region growing and threshold will most likely fail. Model-based methods incorporating anatomical constraints are necessary to separate the renal cortex and renal column. Furthermore, in order to overcome the weak boundaries, graph-based methods can be employed to search for a global optimal solution. In this paper, we propose an extended optimal surface search method [4] to segment the renal

* Corresponding author.

cortex by simultaneously extracting its outer and inner surfaces. The graph construction [5] was utilized to build a single graph for the entire image. However, it could not fully accommodate multiple surfaces, so we propose a new graph construction scheme that models multiple sub-graphs to address this problem.

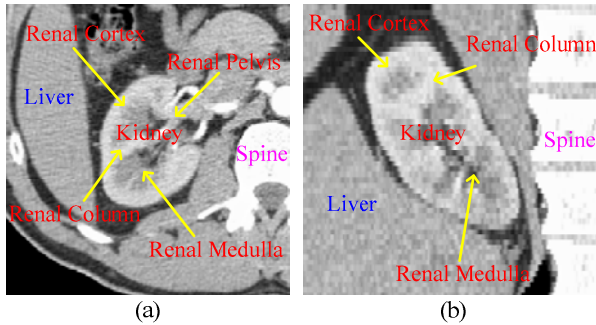


Fig. 1. CT image of the kidney. (a) Coronal view. (b) Sagittal view.

2 Methods

The framework of the proposed approach is illustrated in Fig. 2. The general strategy is to solve the renal cortex segmentation problem through multiple-surfaces extraction. The segmentation has two major stages. The first stage provides the approximate segmentation of the whole kidney, and the second stage simultaneously extracts the two surfaces.

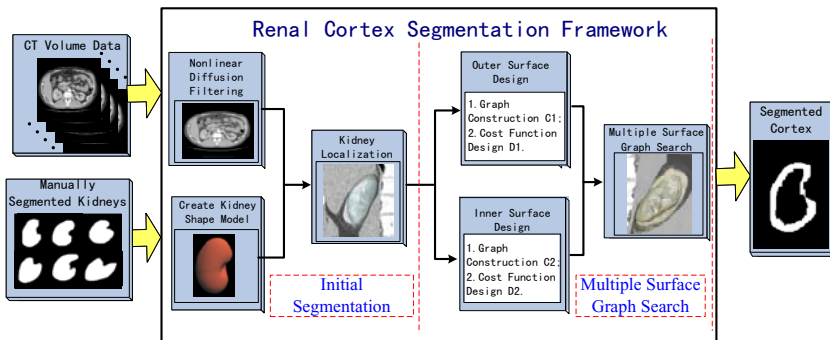


Fig. 2. The proposed renal cortex segmentation framework

2.1 Initial Segmentation

The initial segmentation is based on the statistical shape model method [6-7] and consists of the following three main steps: nonlinear diffusion filtering, kidney shape model creation and kidney initial localization.

1) Nonlinear diffusion filtering

As a preprocessing step, a nonlinear diffusion filter [8] is applied to smooth the image and reduce noise. The diffusion filter is defined by the following equation:

$$\begin{cases} \frac{\partial u(x, y, z, t)}{\partial t} = \text{div}[g(|\nabla u|^2)\nabla u] \\ u(x, y, z, 0) = u_0(x, y, z) \end{cases} \quad (1)$$

t is a scale parameter, and the diffusivity function $g(s) = 1 - \exp(-3.6150/(s/\lambda)^{10})$ if $s > 0$, and $g(s) = 1$ otherwise. λ is the contrast parameter.

2) Kidney shape model creation

The kidney shape model is built from manually segmented kidneys in a training set. First, a triangulated surface mesh is reconstructed from the manually labeled images using marching cubes [9]. Then, a minimum description length (MDL) [10] approach is used to determine the point correspondence for all training shapes. Subsequently, all training shapes are aligned with affine transformations. Finally, the average kidney shape can be obtained with $\bar{\Phi} = \sum_i \Phi_i / n$, where Φ_i is a training surface, and n is the number of the training set.

3) Kidney initial localization

After generating the average kidney shape, we can approximately locate this average shape using the software package Amira [11]. We use the built-in transformation editor to roughly register the average shape to the original volume.

2.2 Multiple-Surface Graph Search

Based on the kidney initialization results, we employ an optimal surface search algorithm [4]. This algorithm can be considered as an optimization process aiming at finding the set of surfaces with the minimum cost. The three major components are graph construction, cost function design and optimal surface recovering.

1) Graph construction

A key innovation of our method is its novel graph construction scheme, enabling construction of different sub-graphs according to different surface properties and inter-surface relationships. The graph construction process is illustrated in Fig. 3. Two sub-graphs (G_0, G_1) are constructed for outer and inner surfaces respectively in a narrowband around the surface meshes. For each vertex v_j on the mesh, a column of equidistant points is sampled along the normal direction of the vertex. The sampled points is denoted as $S_i(v_j^k)$, $i = 0, 1$; $j = 0, \dots, M - 1$; $k = 0, \dots, N_i$ constituting the nodes for the sub-graphs, where i is the index of the two sub-graphs (G_0, G_1), j is the index of the vertex on the surface, and k is the index of the sampled points along the columns. We adopt different strategies to form the columns in sub-graphs G_0 and G_1 , which are written in equ. 2:

$$\begin{cases} S_0(v_j^k) = v_j - (k - (N_0 - 1)/2) \bullet \Delta_0 \bullet \overline{n_j} \quad (k = 0, \dots, N_0 - 1) \\ S_1(v_j^k) = v_j - k \bullet \Delta_1 \bullet \overline{n_j} \quad (k = 0, \dots, N_1 - 1) \end{cases} \quad (2)$$

where the normal direction $\overline{n_j}$ points outward, and (Δ_0, Δ_1) are the smoothness constraints for the two sub-graphs. The outer surface is searched both inward and outward, but the inner surface is searched only inward. This scheme is illustrated in Fig. 3(b)(c) so that G_0 spans both inside and outside of the pre-segmented surface and G_1 spans only inside the surface. This is based on the prior knowledge that the renal cortex inner surface should be inside of the initial exterior surface.

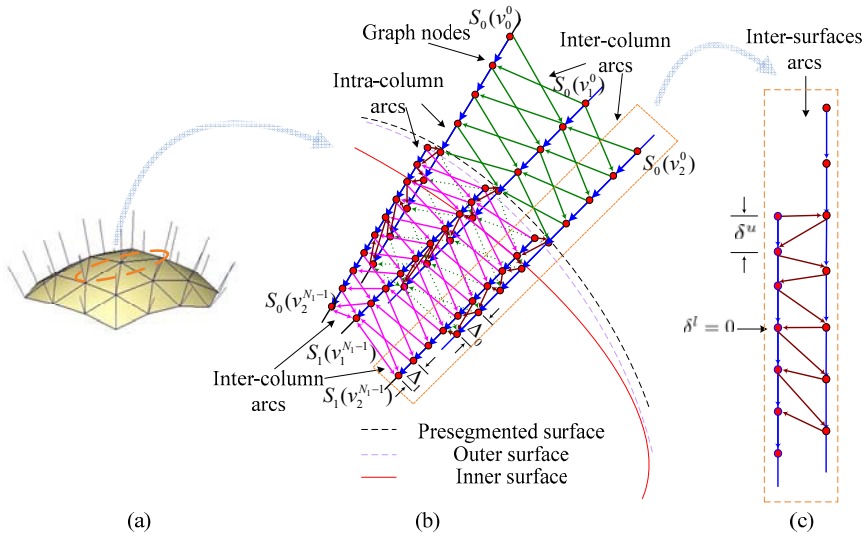


Fig. 3. Graph construction. (a) Triangulated surface mesh with vertex normal. (b) Graph construction with three columns in the orange ellipse in (a). The whole graph is composed of two sub-graphs G_0 and G_1 . The blue edges denote the intra-column arcs, while the green edges and red edges denote the inter-column arcs of the two sub-graph, and the brown edges between two columns projected from the same vertex denote the inter-surface arcs. (c) Detail of the inter-surfaces arcs.

There are three types of arcs in the graph shown in Fig. 3(b)(c):

$$\begin{cases} E^a = \{ \langle S_i(v_j^k), S_i(v_j^{k-1}) \rangle \mid i = 0, 1; j = 0, \dots, M - 1; k = 1, \dots, N_i \} \\ E_i^r = \{ \langle S_i(v_p^k), S_i(v_q^{\max(0, k - \Delta_i)}) \rangle \mid \forall v_p, v_q \text{ is adjacent}; i = 0, 1; k = 1, \dots, N_i \} \\ E^s = \{ \langle S_1(v_i^k), S_0(v_i^k) \rangle, \langle S_0(v_i^k), S_1(v_i^k) \rangle \mid \delta^l < (S_1(v_i^k) - S_0(v_i^k), S_1(v_i^k) - S_0(v_i^k)) < \delta^u \} \end{cases} \quad (3)$$

where (i, j, k) are the indices defined above. The intra-column arc E^a denoted by the blue edges connect two adjacent points on the same column ; the inter-column arcs E_0^r denoted by green edges connect two points on two adjacent columns of the outer surface sub-graph G_0 , while the arcs E_1^r are denoted by red edges in the inner surface sub-graph G_1 ; and the inter-surface arcs E^s denoted by the brown edges connect two points on different sub-graphs. In our graph construction, different smoothness constraints (Δ_0, Δ_1) are designed for the inter-column arcs in the two sub-graphs. Separation constraints (δ^l, δ^u) are exerted on the inter-surface arcs to separate the two surfaces.

2) Cost function design

All of the intra-column, inter-column and inter-surface arcs are viewed as n -links in the s - t graph [12] and assigned infinity values initially. Each node has a weight $W_i(v_j^k)$ in the weighted directed graph, where (i, j, k) are the indices defined in the previous section. The node weight $W_i(v_j^k)$ is defined below:

$$W_i(v_j^k) = \begin{cases} C_i(v_j^k), & k = 0 \\ C_i(v_j^k) - C_i(v_j^{k-1}), & \text{otherwise} \end{cases} \tag{4}$$

where the cost $C_i(v_j^k)$ is defined with cost functions reflecting some properties of the relevant surfaces. $C_0(v_j^k)$ and $C_1(v_j^k)$ represent the two cost functions for sub-graph G_0 and G_1 :

$$\begin{cases} C_0(v_j^k) = -|\nabla I(v_j^k)| \\ C_1(v_j^k) = -F_{sheet}(I(v_j^k)) \end{cases} \tag{5}$$

the 3-D ‘‘sheet filter’’ F_{sheet} [13] is defined as follows:

$$F_{sheet}(I) = \begin{cases} |\lambda_3| \cdot w(\lambda_2; \lambda_3) \cdot w(\lambda_1; \lambda_3) & \lambda_3 < 0 \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

$\lambda_1, \lambda_2, \lambda_3$ ($\lambda_1 \geq \lambda_2 \geq \lambda_3$) are eigenvalues of Hessian matrix $\nabla^2 I$, and w is written as

$$w(\lambda_s; \lambda_t) = \begin{cases} (1 + \lambda_s / |\lambda_t|)^\gamma & \lambda_t \leq \lambda_s \leq 0 \\ (1 - \alpha \cdot \lambda_s / |\lambda_t|)^\gamma & |\lambda_t| / \alpha > \lambda_s > 0 \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

where γ and α are parameters.

If $W_i(v_j^k) \geq 0$, nodes are connected to the sink terminal t by a directed edge with weight $W_i(v_j^k)$, otherwise nodes will be connected to the source terminal s by a directed edge of weight $-W_i(v_j^k)$. These arcs are t -links.

3) Optimal surface recovering

After the above two steps, a s - t graph G_{st} has been derived from the weighted directed graph G . Then we can apply the traditional graph cut algorithm [12] to recover the optimal surfaces. After that, the outer and inner surface meshes are scanned and converted to volumes with the same dimension and spacing as the original CT datasets. Finally, the segmentation of the renal cortex is obtained by subtracting the two volumes.

3 Experiments and Results

The proposed method was tested on 17 clinical CT datasets using Leave-one-Out cross validation. The CT images consist of $512 \times 512 \times 32$ voxels. These data sets were manually segmented by labeling the kidney tissue to create kidney template images. Our method was implemented on a 32-bit desktop PC (2.33 GHz Core 2 and 2 GB RAM) based on the medical imaging toolkit [14] (MITK <http://www.mitk.net/>).

The parameters used in our algorithm are listed in Table 1. The sampling distance varies inversely with the density of local triangulated mesh. Results of the segmented renal cortex as well as the extracted outer and inner surface are shown in Fig. 4. Manual segmentation was used as reference and compared with the automatic segmentation result according to five metrics: true positive volume fraction (TPVF), false positive volume fraction (FPVF), signed relative volume difference (SVD), average symmetric surface distance (D_{Avg}), and average symmetric RMS surface distance (D_{RMS}). These metrics are summarized in Table 2. TPVF is relatively low compared to (1-FPVF), which is due mostly to the cortex segmentation error around the renal pelvis (blue arrow in Fig. 5(b)). Fig. 5(a) shows one example of the surface distance map between an automatic segmentation and a reference segmentation.

Table 1. Parameter values

Stage	Parameter values
Initial segmentation	$\lambda=10$, step size 0.1, iter 30; $n=17$, model radius 0.3, samples number 41, landmarks 2562
Multiple surfaces graph search	Outer surface: $N_o=41$, $\Delta_o=1$, sampling distance 0.2-4.0mm ; Inner surface : $N_i=21$, $\Delta_i=1$, sampling distance 0.1mm ; $M=2562$, $[\delta^l, \delta^u]=[3,8]$ sampling distance unit $\alpha = 0.5$, $\gamma = 0.25$

Table 2. Metrics results (Mean \pm SD)

TPVF(%)	FPVF(%)	SVD (%)	D_{Avg} (mm)	D_{RMS} (mm)
74.10 \pm 3.18	0.08 \pm 1.25	-3.46 \pm 3.33	0.69 \pm 0.20	1.36 \pm 0.29

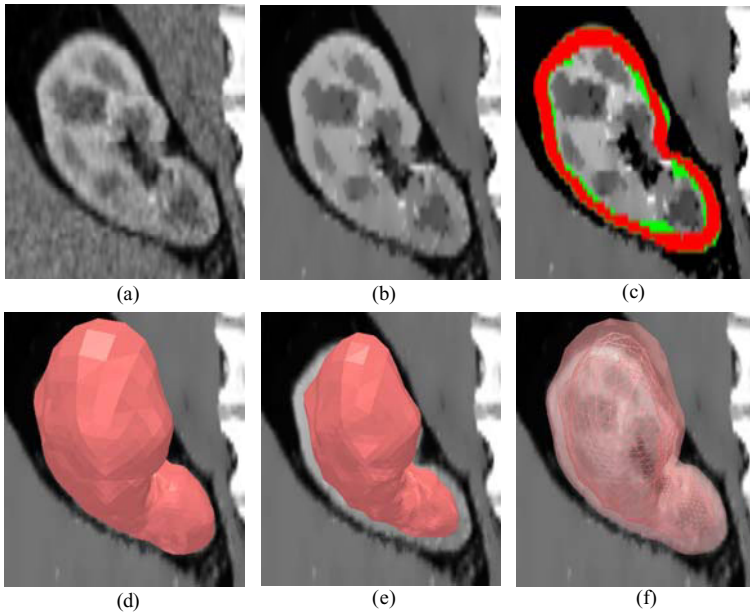


Fig. 4. Surface segmentation results. (a) Original image. (b) Enhanced image using nonlinear diffusion filtering. (c) Red region indicates the renal cortex segmentation result, while the reference segmentation is in green. (d) Outer surface mesh. (e) Inner surface mesh. (f) Combined segmented outer surface (translucent surface) and inner surface (wire frame).

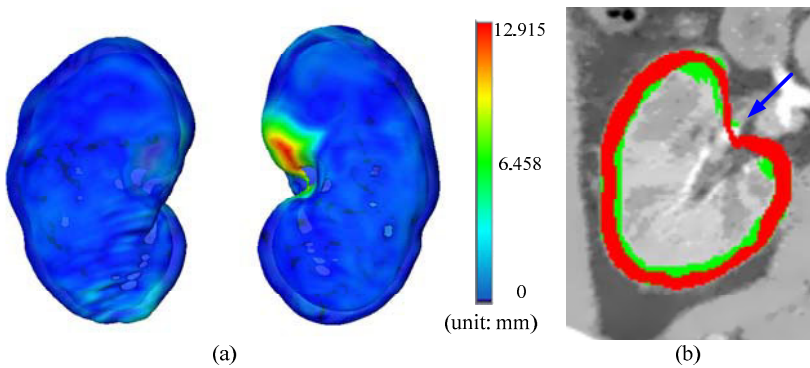


Fig. 5. (a) 3D surface distance map between the outer surface segmentation result and the kidney template image (ground truth). (b) Segmentation error around the renal pelvis (blue arrow). The cortex segmentation is in red, while the reference segmentation is in green.

4 Conclusions and Future Work

A novel approach has been presented to solve the renal cortex segmentation problem. We segment the renal outer and inner surfaces simultaneously by using the optimal

multi-surface search method to generate the renal cortex. The experimental results demonstrate the effectiveness of the proposed method. Our key innovation is the novel graph construction scheme. Different surface sub-graphs are constructed according to the surface properties and inter-surface relationships. This graph construction scheme can be applied in other multiple-surfaces segmentation problems, such as the myocardium and cerebral cortex segmentation.

In our future work, we will focus on improving the performance of inner surface segmentation. More investigations will be conducted about the separation constraints and cost functions, especially the relationship between the two cost functions.

References

1. Lin, D., Lei, C., Hung, S.: Computer-aided kidney segmentation on abdominal CT images. *IEEE T. Inf. Technol. Biomed.* 10(1), 59–65 (2006)
2. Cai, W., et al.: Dynamic-Threshold Level Set Method for Volumetry of Porcine Kidney in CT Images: In Vivo and Ex Vivo Assessment of the Accuracy of Volume Measurement. *Acad. Radiol.* 14(7), 890–896 (2007)
3. Spiegel, M., et al.: Segmentation of kidneys using a new active shape model generation technique based on non-rigid image registration. *Comput. Med. Imaging Graph.* 33(1), 29–39 (2009)
4. Li, K., et al.: Optimal surface segmentation in volumetric images - A graph-theoretic approach. *IEEE Trans. Pattern Anal. Mach. Intell.* 28(1), 119–134 (2006)
5. Yin, Y., et al.: LOGISMOS-Layered Optimal Graph Image Segmentation of Multiple Objects and Surfaces: Cartilage Segmentation in the Knee Joint. *IEEE Trans. Med. Imaging* 29(12), 2023–2037 (2010)
6. Heimann, T., Meinzer, H.: Statistical shape models for 3D medical image segmentation: A review. *Med. Image Anal.* 13, 543–563 (2009)
7. Zhang, X., et al.: Automatic Liver Segmentation Using a Statistical Shape Model with Optimal Surface Detection. *IEEE Trans. Biomed. Eng.* 57(10), 2622–2626 (2010)
8. Weickert, J., Romeny, B., Viergever, M.: Efficient and reliable schemes for nonlinear diffusion filtering. *IEEE Trans. Image Process.* 7(3), 398–410 (1998)
9. Lorensen, W.E., Cline, H.E.: Marching cubes: A high resolution 3D surface construction algorithm. In: *SIGGRAPH 1987 Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques*, pp. 163–170 (1987)
10. Heimann, T., et al.: Implementing the automatic generation of 3d statistical shape models with itk. In: *MICCAI 2006 Open Source Workshop*, Copenhagen, Denmark (2006)
11. Stalling, D., Westerhoff, M., Hege, H.: Amira: A highly interactive system for visual data analysis. In: *The Visualization Handbook*. Elsevier, Amsterdam (2005)
12. Boykov, Y., Kolmogorov, V.: An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Trans. Pattern Anal. Mach. Intell.* 26(9), 1124–1137 (2004)
13. Sato, Y., et al.: Tissue classification based on 3D local intensity structures for volume rendering. *IEEE Trans. Vis. Comput. Graph.* 6(2), 160–180 (2002)
14. Tian, J., Xue, J., Dai, Y., et al.: A novel software platform for medical image processing and analyzing. *IEEE T. Inf. Technol. Biomed.* 12, 800–812 (2008)

3D Modeling of Coronary Artery Bifurcations from CTA and Conventional Coronary Angiography

Rubén Cárdenes¹, Jose Luis Díez², Ignacio Larrabide¹,
Hrvoje Bogunović¹, and Alejandro F. Frangi¹

¹ Center for Computational Imaging & Simulation Technologies in Biomedicine,
(CISTIB) – Universitat Pompeu Fabra and CIBER-BBN, Barcelona, Spain

² Cardiology Department, Dr. Peset Hospital, Valencia, Spain

Abstract. Coronary artery bifurcations are regions where the atherosclerotic plaque appears more frequently and where the percutaneous treatment is more challenging. To analyze these important vascular regions, in this paper is proposed a method for the extraction of realistic 3D models of coronary bifurcations combining information from pre operative computer tomography angiography (CTA) to obtain the 3D structure of the vessels and pre and post operative conventional coronary angiography (CCA) to extract a more accurate estimation of the lumen radius before and after stenting. The method proposed is semiautomatic, starting from a set of user defined landmarks, and has been successfully applied to data from five patients that underwent endovascular treatment in a coronary bifurcation. The results obtained are satisfactory by visual inspection and in comparison with manual measurements.

1 Introduction and Background

Coronary artery bifurcations are places where atherosclerotic plaque formation occurs more frequently, and are challenging vascular regions for percutaneous treatment. Among the treatment techniques described for bifurcations, the one more appropriate for an individual case is still cause of controversy among endovascular interventionists. If a realistic 3D geometry of the bifurcation before and after treatment is available, a detailed analysis of the vascular morphology and its changes could respond to questions such as the level of success of the treatment employed or, in a large population study, what are the techniques more suitable for specific morphologies and locations. Also, computational fluid dynamics studies could be carried out using these geometries, and the study of the changes in hemodynamic parameters, could reveal regions of high risk of re-stenosis.

Several approaches have been proposed to obtain accurate segmentations of coronary arteries. Some important ones are minimum cost path techniques like the one presented in [1] that used a 4D representation of curves to obtain tubular structures from CT and MR data. Another similar approach was proposed in [2]

to extract the vessel centerlines and then the vessel surface using fast marching. However, these and many other approaches are based on information about only one image modality, usually computer tomography angiography (CTA), and conventional coronary angiography (CCA) is not commonly used. In this paper we propose to combine these two modalities, taking advantage of each one to overcome the limitations of the other. Therefore, realistic 3D bifurcation geometries can be obtained using routine acquisition imaging modalities, with a semiautomatic method that provides the pre and post operative models of the vessels. This is, as far as we know, the first time that these two modalities are combined for reconstruction of coronary artery bifurcations.

2 Method

The method proposed here consists of three steps. First, the 3D structure of the bifurcation is obtained from CTA pre operative data using a ridge extraction method followed by a fast marching (FM) approach. After that, CCA images obtained before and after endovascular treatment are processed to obtain the vessel radii along the vessel centerlines, and finally, the models are constructed mapping the vessel radii to the 3D structure.

2.1 3D Centerline Extraction

The first stage of the method uses the CTA image to extract the information of the vessel centerlines forming the bifurcation. In this step, the user specifies the center of the bifurcation in the CTA image, that will be denoted as \mathbf{b}^{3D} , and a minimum of three points on the vessels forming the bifurcation, denoted as \mathbf{p}_1^{3D} for the proximal segment of the main artery, \mathbf{p}_2^{3D} , for the distal part of the main artery and \mathbf{p}_3^{3D} for the secondary artery. Then, a ridge extraction is performed over the image using \mathbf{p}_1^{3D} , \mathbf{p}_2^{3D} , and \mathbf{p}_3^{3D} as seeds, to get the centerlines [3]. To obtain better results, the vessel enhancing filtering by [4] is applied to the CTA image prior to the ridge extraction.

The centerlines obtained from the ridge extraction process do not always form a connected set. Two segments are considered disconnected if they are separated by a distance greater than the step size used in the ridge extraction, (here the step size used is 0.3 mm, lower than the minimum spatial resolution of the CTA, 0.445 mm). These disconnections are due to several reasons, such as the relatively low spatial and temporal resolution of the CTA image, the noise, and especially due to lesions affecting the bifurcation. For this reason, a reconnection strategy is employed similarly to [2]. The method consists in finding for each disconnected segment, the closest point in any of the other segments, \mathbf{y} , using a distance map $u(\mathbf{x})$. Then, a minimum cost path is computed from \mathbf{y} to the initial segment using $u(\mathbf{x})$. To compute $u(\mathbf{x})$ a FM algorithm is employed, solving the Eikonal equation:

$$|\nabla u(\mathbf{x})|F(\mathbf{x}) = 1, \quad (1)$$

and using the disconnected segment points as seeds. The FM computation is stopped when any of the other segment points, \mathbf{y} , is reached. Here, the speed function $F(\mathbf{x})$ has to be carefully selected. A constant speed function will provide an Euclidean distance map, that will connect the vessel centerlines by straight lines, thus, providing non realistic results (see green points in Fig. 1). In order to follow a more realistic trajectory of the vessels, we use a speed function that depends exponentially (as in [5]) on the local vessel structure as follows:

$$F(\mathbf{x}) = e^{(\alpha \cdot V(\mathbf{x})/M)}, \quad (2)$$

where $V(\mathbf{x})$ is the multiscale vessellness of the CTA, computed as in [6], and M is the maximum value in $V(\mathbf{x})$. Ten scales are used to compute $V(\mathbf{x})$, from 0.2 up to 4 mm. $F(\mathbf{x})$ is designed to drive the FM process faster along vessels and the parameter α controls its strength. Low α values give a speed closer to constant everywhere, and high values increment the speed at high vessellness regions. Here we use $\alpha = 2$ that provides good results in our experiments. The reconnection process is repeated for each segment until there is only one connected segment. The behavior of this reconnection strategy is shown in the phantom shown in Fig. 1 where three disconnected segments are connected following the vessel structures.

In Fig. 2 the centerline points obtained from the ridge extraction (red points) and reconnection step (green points) are shown for a left coronary artery tree. To take into account the calcified tissues of the plaque, that if present, are visible in the CTA, two different centerlines are computed for the pre and post operative models. For the post operative centerline, the ridge extraction is computed using the original CTA image. However, for the pre operative centerline, the calcifications are first subtracted from the CTA image using a threshold of 721 Hounsfield units [7], and then, the centerlines are computed again. Therefore, the pre operative centerlines deviate with respect to the previous ones at calcified lesions as shown in Fig. 2.

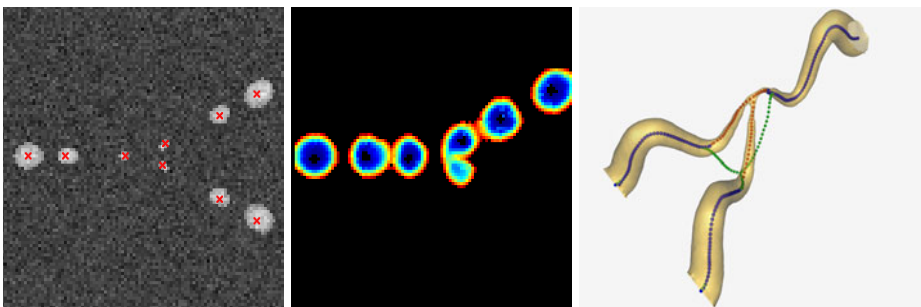


Fig. 1. Slice of a digital phantom with centerline points overlaid (left), same slice with $u(\mathbf{x})$ computed from one segment (middle), and 3D view of the phantom (right). Initial points are in blue, reconnected points in red, and in green connected points using $F(\mathbf{x}) = 1$.

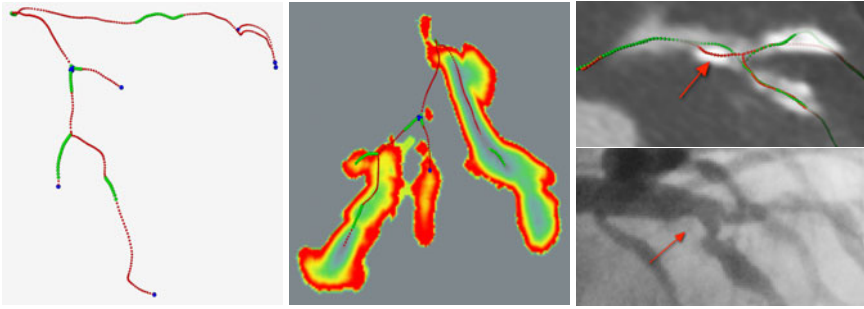


Fig. 2. Left: Initial centerline points (red) obtained from coronary CTA, connected centerline points using FM (green) and end points and bifurcation points (blue). Middle: slice of the distance field $u(\mathbf{x})$ from one segment. Right top: centerlines of the post operative model (red), and pre operative centerline (green) avoiding the calcified plaque pointed by arrows. Right bottom: corresponding CCA image.

2.2 Processing of the 2D Angiographic Images

In the following step, the vessel radii on the CCA images are calculated. In particular, two images are selected from pre and post operative views, using an angle where the area of the projection of the bifurcation is maximum. This projection corresponds approximately to the CCA image chosen by the clinician for an optimum visualization of the bifurcation, minimizing foreshortening and vessel overlap. Notice that, only vessel radii and lengths from the bifurcation center are used from CCA, and therefore there is no need to have a perfect projection matching between the CTA and the CCA. This is especially important because the vessel geometry extracted from CTA can vary with respect to the geometry extracted from CCA when the guide wire is inserted, that could change the vessels curvature (see Fig. 3). Therefore, our method does not need perfect

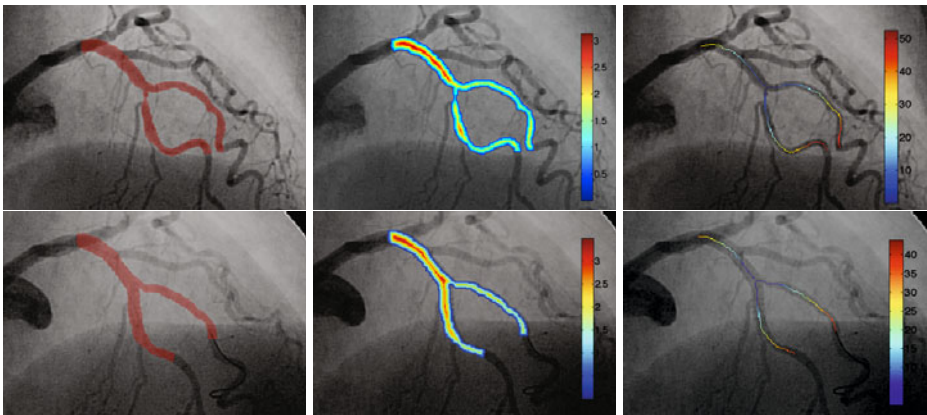


Fig. 3. CCA images. Top row: pre operative, bottom row: post operative. Overlay images are from left to right: segmentation, distances from the vessel boundary, and centerline colored with the lengths to b^{2D} .

correspondence between CTA and CCA, being robust with respect to geometry changes and with respect to the cardiac phase used in both modalities, although it is chosen to be approximately the same. One alternative method could be to perform an elastic 2D registration between CCA and CTA projections to find radii correspondences, but the lack of information about the projection geometry of the CCA acquisition would increase considerably the complexity of the problem as well as the computational cost.

For each of the two projection images selected, the bifurcation center \mathbf{b}^{2D} is also manually selected, corresponding to the point \mathbf{b}^{3D} . Also, three points on the branches are marked (\mathbf{p}_1^{2D} , \mathbf{p}_2^{2D} and \mathbf{p}_3^{2D}), that will be used to assign correspondences between the 2D and 3D centerlines and to stop the 2D segmentation process, but that do not have to correspond to \mathbf{p}_i^{3D} . The lumen in the CCA images are first segmented, then the vessel centerlines are computed, and finally the vessel radii are computed as the distances from the segmented boundary at the centerline points. For the segmentation, a level set approach is employed starting from \mathbf{b}^{2D} and is stopped when it reaches the user defined points \mathbf{p}_i^{2D} . Two pre and post operative CCA images are shown in Fig. 3 with the segmentation, the distances from the segmented boundaries and the centerlines overlaid.

2.3 Radius Mapping

In the next step the radii values are mapped from 2D to 3D, to construct the bifurcation models. Here, circular cross-section can be assumed for the arteries without losing too much accuracy in the majority of cases, because this assumption only fails in 20% to 30% of the lesions [8].

The radius mapping is implemented using the centerline lengths. Let $\Gamma_i(s)$ and $\gamma_i(s)$ be the parameterized centerline curves of each vessel in 3D and 2D respectively, where $\Gamma_i(0) = \mathbf{b}^{3D}$, $\Gamma_i(1) = \mathbf{p}_i^{3D}$ and $\gamma_i(0) = \mathbf{b}^{2D}$. Then, we choose $\gamma_i(1)$ as the 2D point such as corresponding curves have the same length, $L(\Gamma_i(1)) = L(\gamma_i(1))$, where the length function $L(\gamma(s))$ is given by

$$L(\gamma(s)) = \int_0^s |\dot{\gamma}(t)| \gamma(t) dt . \quad (3)$$

Then, a radius value is assigned to each point of Γ_i , using the radii values computed previously in the CCA images. The mapping is performed using distances from the bifurcation center, such that if $R_i(d)$ and $r_i(d)$, are the radii in 3D and 2D at distance d from \mathbf{b}^{3D} and \mathbf{b}^{2D} respectively, at the i^{th} branch, then, the radius mapping can be expressed as $R_i(L(\Gamma_i(s))) = r_i(L(\gamma_i(s)))$, $\forall s \in [0, 1]$.

In general, there is no exact correspondence between the points $\Gamma_i(s)$ and $\gamma_i(s)$ due to foreshortening. However, we can assume that in a small region near \mathbf{b}^{3D} , the vessel centerlines lay near a plane [9], and thus, the foreshortening effect is minimum on this plane and the radius mapping can be applied. Finally, the obtained surfaces are smoothed to avoid sharp transitions at vessel junctions.

3 Results

We have selected 5 patients for this study, affected by occlusion in a coronary bifurcation and eligible for percutaneous intervention. All the patients had a pre operative CTA scan, which was acquired in a 64 multi-slice scanner from GE Medical Systems, with an image resolution of $0.445 \times 0.445 \times 0.625$ mm, and at least two sequences of pre and post operative CCAs as described above.

In order to show our results, we have selected for each case, the CTA oblique slice that approximates best to the plane formed locally by the bifurcation vessels, that also fits best with the CCAs used, and will be denoted as bifurcation plane. This plane is shown in Fig. 5 (left) for one case, that corresponds to the tangent plane of the epicardium. Fig. 4 illustrates the bifurcation region on each CCA (top row), the CTA bifurcation plane with the models and a manual segmentation outlines overlaid (middle row), and the pre and post operative bifurcation models for the five patients (bottom row). The differences between the two models and the matching with the real images are clearly seen. In the first and fifth cases, the secondary branch is bigger in the pre operative model (green color inside), which is in accordance with the diameters observed in the CCAs, and could be due to partial occlusion of the secondary branch after stenting.

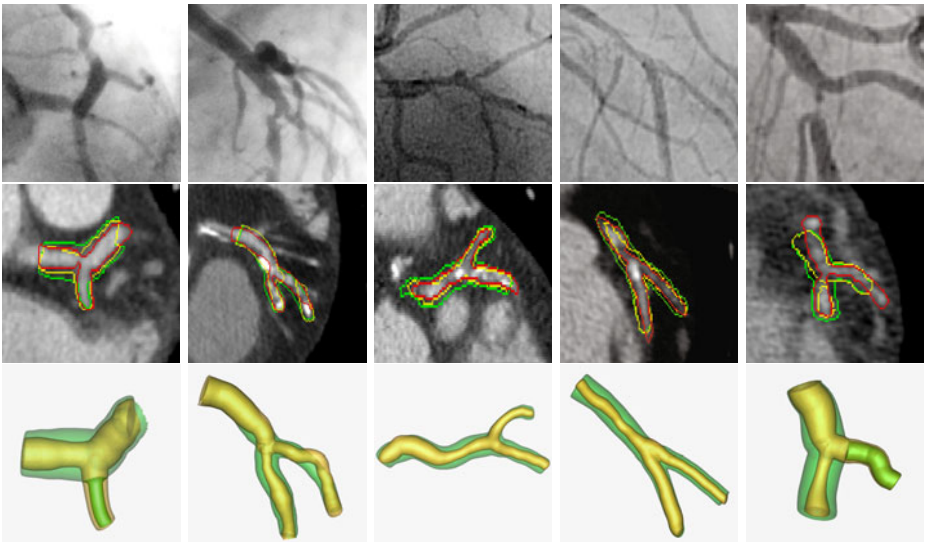


Fig. 4. Detail of the CCAs used (top row), corresponding bifurcation plane from the CTA with manual segmentation (red), pre (yellow) and post operative (green) outlines overlaid (middle row), and 3D bifurcation models (bottom row); in yellow and green the pre and post operative models respectively.

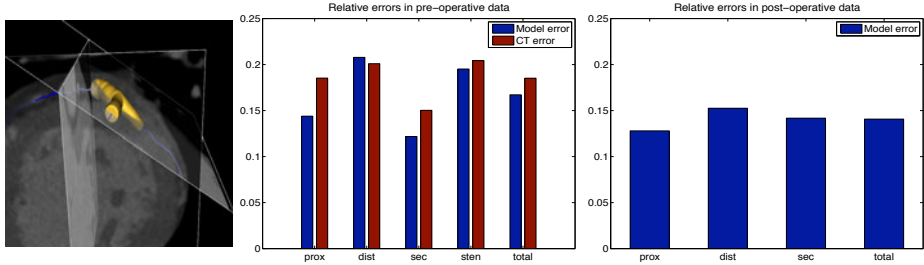


Fig. 5. Bifurcation plane in one case (left), and relative errors obtained in the diameters measures for the pre operative data (middle) and for the post operative data (right)

4 Comparison with Manual Measurements

To evaluate our results, two comparisons have been carried out. First, manual segmentation is performed in the bifurcation plane extracted from the CTA, and is compared with the pre operative model at the same plane, computing the dice similarity index between the two regions, whose boundaries are shown in Fig. 4 (middle row). The values obtained are shown in table 1, where a good agreement is found (0.7675 in average).

A second comparison has been done measuring the three vessels diameters in the CTA, the CCAs and in our models using an annotation tool. The diameter at the stenosis is also measured in all the pre operative data. Then, considering the diameters from the CCA as the ground truth, the relative errors of our method and from CTA are computed for the pre operative data, see Fig. 5 (middle). These errors are lower in our models than in CTA, showing the importance of including the radius information from CCA. The errors obtained between our post operative models and the post operative CCAs are also plotted in Fig. 5 (right), where the values obtained are slightly smaller than before.

Table 1. Dice similarity index computed between manual delineations in the bifurcation plane and a slice of the same plane of the pre operative models for the 5 patients.

	patient 1	patient 2	patient 3	patient 4	patient 5	average
DS :	0.8171	0.7217	0.8244	0.8197	0.6545	0.7675

5 Conclusions

We have proposed a method for realistic modeling of coronary bifurcations combining information from two modalities: CTA and CCA. We have shown the advantages of using both modalities together: from CTA, the 3D vessel trajectories and the visible atherosclerotic plaque are used, and from CCA we obtain a more accurate estimation of the lumen radius, before and after stenting. The results presented from five patients are successful, and the models obtained match

well with both the CTA and the CCA data as shown above. Numerical comparison with manual segmentation shows a good accuracy. Also, from comparisons of the diameters we have also shown that this method can improve the accuracy with respect to CTA used alone. Therefore, using images acquired during normal clinical practice, it is possible to obtain accurate models from specific regions of the coronary artery tree, before and after percutaneous intervention. These accurate models can be extremely useful to assess geometrically the outcome of endovascular interventions and they can be obtained very efficiently.

Two assumptions have been used: circular cross section of the vessels, that is not true only in 20% to 30% of the lesions, and planar correspondence between the CCAs and the bifurcation vessels in a small region around its center. From the results obtained, it is shown qualitatively and quantitatively, that both approximations are valid with the data sets used here, and therefore this method shows to be a good technique to obtain realistic results. However, a wider study with a larger population, analyzing user input variability, cardiac phase selection and acquisition plane selection, is still needed to fully validate this method.

Acknowledgements. This work has been funded by a CDTI CENIT-cvREMOD grant of the Spanish Ministry of Science and Innovation, R.C. is funded by a Beatriu de Pinós fellowship from AGAUR, and A.F.F. is partially funded by the ICREA-Academia programme, from the Regional Government of Catalonia.

References

1. Li, H., Yezzi, A.: Vessels as 4-D curves: Global minimal 4-D paths to extract 3-D tubular surfaces and centerlines. *IEEE T. Med. Imag.* 26(9), 1213–1223 (2007)
2. Mueller, D., Maeder, A.: Robust semi-automated path extraction for visualising stenosis of the coronary arteries. *Comput. Med. Imag. Grap.* 32, 463–475 (2008)
3. Aylward, S., Pace, D., et al.: TubeTK, segmentation, registration, and analysis of tubular structures in images. Kitware, Inc. (2010)
4. Manniesing, R., Viergever, M.A., Niessen, W.J.: Vessel enhancing diffusion: A scale space representation of vessel structures. *Med. Image Anal.* 10(6), 815–825 (2006)
5. Hassouna, M.S., Farag, A.A.: Variational curve skeletons using gradient vector flow. *IEEE T. Pattern Anal.* 31(12), 2257–2274 (2009)
6. Frangi, A.F., Niessen, W.J., Vincken, K.L., Viergever, M.A.: Multiscale vessel enhancement filtering. In: Wells, W.M., Colchester, A., Delp, S.L. (eds.) *MICCAI 1998*. LNCS, vol. 1496, pp. 130–137. Springer, Heidelberg (1998)
7. Sakakura, K., Yasu, T., Kobayashi, Y., Katayama, T., Sugawara, Y., Funayama, H., Takagi, Y., Ikeda, N., Ishida, T., Tsuruya, Y., Kubo, N., Saito, M.: Noninvasive tissue characterization of coronary arterial plaque by 16-slice CT in acute coronary syndrome. *Angiology* 52(2), 155–160 (2006)
8. Arbab-Zadeh, A., Texter, J., Ostbye, K.: Quantification of lumen stenoses with known dimensions by conventional angiography and CT? implications for the use of angiography as a gold standard. *Heart* 96, 1358–1363 (2010)
9. Fung, Y.: *Biomechanics*, 2nd edn. Springer, New York (1997)

Efficient Detection of Native and Bypass Coronary Ostia in Cardiac CT Volumes: Anatomical vs. Pathological Structures

Yefeng Zheng¹, Huseyin Tek¹, Gareth Funka-Lea¹,
S. Kevin Zhou¹, Fernando Vega-Higuera², and Dorin Comaniciu¹

¹ Image Analytics & Informatics, Siemens Corporate Research, Princeton, NJ, USA

² Computed Tomography, Healthcare Sector, Siemens AG, Forchheim, Germany
yefeng.zheng@siemens.com

Abstract. Cardiac computed tomography (CT) is the primary non-invasive imaging modality to diagnose coronary artery disease. Though various methods have been proposed for coronary artery segmentation, most rely on at least one user click to provide a seed point for initialization. Automatic detection of the coronary ostia (where coronaries originate from the aorta), including both the native coronary ostia and graft ostia of the bypass coronaries, can make the whole coronary exam workflow fully automatic, therefore increasing a physician's throughput. Anatomical structures (native coronary ostia) and pathological structures (graft ostia) often require significantly different detection methods. The native coronary ostia are well constrained by the surrounding structures, therefore are detected as a global object. Detecting the graft ostia is far more difficult due to the large variation in graft position. A new searching strategy is proposed to efficiently guide the focus of analysis and, at the same time, reduce the false positive detections. Since the bypass coronaries are grafted on the ascending aorta surface, the ascending aorta is first segmented to constrain the search. The quantitative prior distribution of the graft ostia on the aorta surface is learned from a training set to significantly reduce the searching space further. Efficient local image features are extracted around each candidate point on the aorta surface to train a detector. The proposed method is computationally efficient, taking about 0.40 seconds to detect both native and graft ostia in a volume with around $512 \times 512 \times 200$ voxels.

1 Introduction

Cardiovascular disease (CVD) is the first leading cause of death in the United States and coronary stenosis (narrowing of the vessel) is the most common CVD [1]. If the stenosis is too severe and medical therapy does not help, an artery or vein from elsewhere in the patient's body is often harvested and grafted to the coronary arteries to bypass the stenosis (see Fig. 1). Cardiac computed tomography (CT) is the primary non-invasive imaging modality to diagnose coronary stenosis thanks to its superior image resolution. To facilitate the diagnosis, it

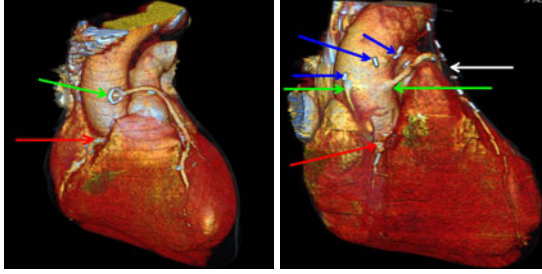


Fig. 1. 3D coronary volumetric visualization for two datasets, showing the right coronary ostia (red arrows), graft ostia of bypass coronaries (green arrows), left internal thoracic artery bypass (white arrow), and surgical clips (blue arrows). The left coronary ostium is not visible due to occlusion of the right ventricle.

is very important to develop a robust system that can efficiently extract the coronary artery centerline, followed by vessel lumen segmentation, to provide quantification of the coronary stenosis (i.e., measuring the percentage of the lumen area blocked by plaques). Though various methods have been proposed for coronary artery segmentation, most rely on at least one user click to provide a seed point to initialize the centerline extraction [2]. Coronary segmentation is a time consuming task. It is desirable to perform it automatically in a preprocessing step, therefore, when a physician starts the exam, the segmentation is readily available. Automatic detection of the coronary ostia opens the potential to make the whole workflow fully automatic, therefore increasing a physician's throughput.

Almost all previous methods on native coronary ostium detection start with an explicit segmentation of the aorta first. The coronary artery is detected as a tubular structure attached to the aorta and the position of the attachment is taken as a detected ostium. Many methods require at least one user click inside the aorta to initialize the aorta segmentation, even though one click is not sufficient to guarantee good detection of the ostia [3]. Recently a few automatic coronary ostium detection methods have been proposed. Wang and Smedby [4] used a Hough transform based 2D circle detection algorithm to detect and trace the ascending aorta. The largest connected components on each side of the aortic root were picked as the left and right coronary arteries. The method proposed by Tek *et al.* [5] also performed automatic aorta segmentation. They then tried to trace coronary centerlines from the aorta surface in order to detect coronary ostia. Since the computationally expensive centerline tracing algorithm needed to be run on the whole aorta surface, their approach was slow. Some native coronary ostium detection methods can be extended to detect graft ostia. However, the performance is either not reported [5,4] or poor [3].

As a prerequisite for all the following coronary analysis procedures, coronary ostium detection have to be fast and robust. In this paper, we propose an efficient method to automatically detect both native coronary ostia and graft ostia of bypass coronaries. Anatomical structures (native coronary ostia) and pathological

structures (graft ostia) often require significantly different detection methods. Since the native coronary ostia are well constrained by the neighboring global structures, we can detect them directly without explicit segmentation of the aorta as done in the previous work [3,4,5]. As a pathological structure, a graft ostium has far less constraint in its exact position (see Fig. 1). The number of bypass coronaries also varies quite a lot, from one up to four. To address these challenges, a different searching strategy is proposed to efficiently guide the focus of analysis and, at the same time, reduce the false positive detections. We first segment the ascending aorta and constrain the graft ostium detection onto the aorta surface where a bypass coronary is anastomosed.¹ The quantitative prior distribution of the graft ostia on the aorta surface is learned from a training set to significantly reduce the searching space further. Efficient local image features are extracted around each candidate point on the aorta surface to train a detector. The proposed approach is computationally efficient, taking about 0.40 seconds to detect both native and graft ostia.

2 Native Coronary Ostium Detection

As an anatomical structure, the native coronary ostium has strong constraints that facilitate the automatic detection. For example, a patient contains two (and only two) native coronary ostia and the left/right ostium is located on the left/right cusp of the aortic root, respectively. For chronic total occlusion patients, the artery originated from a coronary ostium may be completely obstructed thus not visible in a cardiac CT volume. However, the correct position of the native ostium can still be inferred from the surrounding tissues. In this paper, we propose an efficient (taking a fraction of a second) and robust method to detect native coronary ostia as a global object. Different to the previous methods [3,5,4], no explicit aorta segmentation is necessary. There are two coronary ostia. Though it is possible to detect each independently, the detection results may be inconsistent in geometry. It also wastes computation power by ignoring the strong geometric constraint between the ostia. We propose to detect them as a global object (a bounding box) comprised with both coronary ostia. From the position, orientation, and scale of this global object, we can infer the position of individual ostium. As shown in Fig. 2, the bounding box is defined as a cube with one side aligned with the direction connecting the left and right coronary ostia, and the second side aligned with the aortic root centerline. The coronary ostia are located at the center of two opposite faces of the cube. The marginal space learning (MSL) method [6] is used to detect the bounding box. MSL is an efficient and robust method to detect a 3D anatomical structure in medical images based on discriminative machine learning techniques. Due to the space limit, we would like to refer readers to [6] for more details of MSL.

¹ We do not need to detect the the left/right internal thoracic artery (LITA/RITA) ostium of the in-situ LITA/RITA graft since the standard cardiac CT protocol does not cover the ostium on the subclavian artery (as shown in Fig. 1).

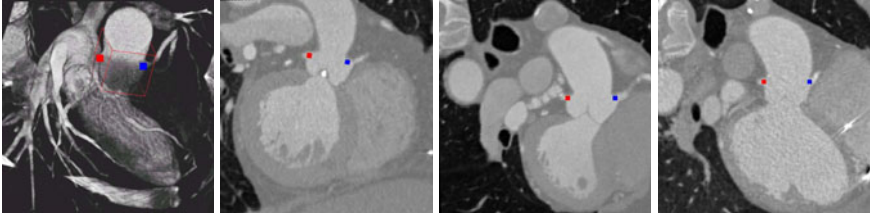


Fig. 2. Native coronary ostium detection. First column shows the detected bounding box of the coronary ostia. The right columns show the detection results on three datasets (red dot for the left ostium and blue for the right).

3 Graft Ostium Detection for Bypass Coronaries

3.1 Learning Prior Distribution of Graft Ostia

The bypass coronary arteries are attached on the ascending aorta. We can reduce the detection errors by constraining the searching of graft ostia onto the aorta surface, which is segmented automatically using the method proposed in [7]. Though distribution of the graft ostia on the aorta surface is widely spread, a bypass coronary is always grafted on the anterior surface of the ascending aorta (as shown in Fig. 3) since if it is grafted in a posterior position, a physician needs to wind the bypass coronary around the aorta to bring it to the heart surface, which introduces unnecessary risks associated with the pressure caused by motion of the aorta and heart chambers. However, such observation is vague domain knowledge and a quantitative prior distribution of the graft ostia is desired to optimally use this domain knowledge.

To learn the prior distribution probability of graft ostia, we need to establish the point correspondence of the aorta surface meshes across different volumes. The length of the aorta captured in a volume varies significantly, presenting a big challenge to volume registration and surface mesh registration methods. Furthermore, the aorta has a simple tubular structure with a smooth surface. It is difficult to identify a fiducial point to establish the correspondence. In this paper, we introduce a consistent resampling method to establish the mesh correspondence by exploiting the convenience of the tubular shape of the aorta. First, we uniformly resample the aorta centerline under a certain resolution (1 mm in our case), starting from the aortic root and extending to the aortic arch. In this way, we establish the correspondence of the aorta centerline after resampling, even though its length varies. At each resampled centerline point, we calculate the 2D intersection of the aorta mesh and the plane perpendicular to the centerline. The 2D intersection contour is close to an ellipse, which is then uniformly resampled to 96 points with a clock-wise ordering. To establish a consistent ordering of these 96 points, the point with the smallest x -coordinate position is selected as the first point. The number of sampling points (96 in this case) is selected such that the distance between neighboring resampled points is close to 1 mm. Thus, we establish the point correspondence of the 2D mesh

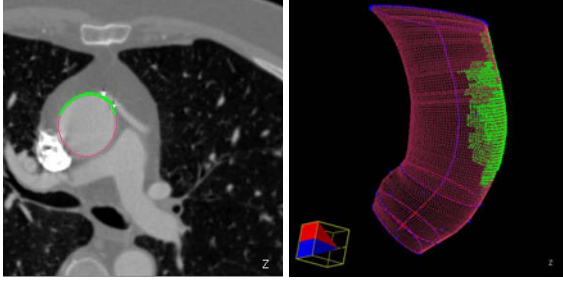


Fig. 3. Constrained searching range of the graft ostia of bypass coronaries on the aorta surface. **Left:** The graft ostia are constrained to the anterior part of the ascending aorta. **Right:** The distribution of graft ostia on the 3D aorta surface mesh. Note: The aorta mesh is shown in red, while the green dots show the constrained searching region for the graft ostia.

intersection. After triangulating the resampled points into a surface mesh (as shown in Fig. 3), we establish good correspondence for the aorta mesh points across different volumes, therefore we can align all meshes into a common coordinate system.

Given a training set of M volumes with annotated graft ostia, we use the Parzen window based method to estimate the prior distribution density. Here, the Gaussian kernel is used

$$P_0(x) = \frac{1}{M} \sum_{i=1}^M \left[\sum_{j=1}^{N_i} \frac{1}{2\pi\sigma} \exp\left(-\frac{\|x - x_i^j\|^2}{2\sigma^2}\right) \right], \quad (1)$$

where N_i is the number of graft ostia in training volume i , x_i^j is the position of j^{th} graft ostium in volume i , and σ is the standard deviation of the Gaussian kernel. Throughout our experiments, we set $\sigma = 1$ mm and the Gaussian distribution is truncated after 3σ .

After learning prior distribution of graft ostia from a training set, only mesh points with $P_0(x) > 0$ (green dots in Fig. 3) are considered in the following graft ostium detection. Those points with $P_0(x) = 0$ can be excluded. Depending on the portion of the aorta captured in a CT volume, the excluded region varies. In our datasets, on average 80% of the aorta surface can be safely excluded. The prior probability $P_0(x)$ at each mesh point can be further used to weight the detection score during graft ostium detection, as discussed in the next section.

3.2 Graft Ostium Detection

We train a classifier that can tell us whether a mesh point is a graft ostium or not. Since the aorta mesh is so densely resampled, it often contains more than 10,000 points. After excluding all points with $P_0(x) = 0$ using the prior distribution, we still need to classify a relatively large number of samples ($\sim 2,000$ points). Therefore, the image features used for classification need to be efficient. In this

paper, we use the steerable features as proposed in [6]. As suggested by [6], we use a regular sampling pattern with $5 \times 5 \times 5$ sampling points. The sampling pattern is aligned with the aorta surface normal at the mesh point under testing. At each sampling point, we extract 24 local image features as specified in [6]. In total, we get $5 \times 5 \times 5 \times 24 = 3000$ features. The extracted features are used to train a boosting classifier [8], which outputs a classification score in the range of $[0, 1]$. The higher the score is, the more likely the candidate is a graft ostium.

We want to integrate the prior distribution of a graft ostium with the classification score to further improve the detection accuracy. Suppose $E = 1$ represents the event that a graft ostium presents and $F(x)$ is the feature vector extracted at mesh point x . We approximate the posterior probability $P(E = 1|x, F(x))$ of observing a graft ostium at position x as

$$P(E = 1|x, F(x)) \propto P(E = 1|x)P(E = 1|F(x)), \quad (2)$$

where $P(E = 1|x)$ is the prior probability of observing a graft ostium at x and $P(E = 1|F(x))$ is the conditional probability given feature vector $F(x)$. Here, $P(E = 1|F(x))$ is replaced by the classification score, which is a good approximate as shown in [8]. In practice, we normalize $P(E = 1|x)$ with its maximum value

$$P(E = 1|x, F(x)) = \frac{P(E = 1|x)}{\max_x \{P(E = 1|x)\}} P(E = 1|F(x)). \quad (3)$$

After classification, all points with an estimated posterior probability larger than 0.5 are kept. We then perform clustering analysis on the aorta surface. The cluster center is output as a detected graft ostium. Additional validation (e.g., coronary centerline tracing from detected ostium) can further reduce the false positives.

4 Experiments

To evaluate the accuracy of the native coronary ostium detection, we collected and annotated 1360 datasets. A four-fold cross-validation is performed for evaluation. The Euclidean distance from the detected left coronary ostium to the corresponding ground truth has a mean of 1.66 mm (standard deviation of 1.15 mm) and a median of 1.45 mm. The corresponding errors of the right coronary ostium are 1.73 mm for the mean (standard deviation of 1.71 mm) and 1.31 mm for the median. The maximum detection error of the left and right coronary ostia is 9.05 mm. Only about 0.8% of datasets have a detection error larger than 5 mm (which is approximately the diameter of the coronary arteries around the ostia) and are treated as failures. Fig. 2 shows some examples of the detected native coronary ostia. We also compare with the method proposed by Tek *et al.* [5]. Their success rate on these datasets is about 90%, which is significantly lower than ours.

Among the 1360 datasets, 224 have bypass coronaries with a total of 367 graft ostia. Again, a four-fold cross-validation is performed for evaluation. It is not trivial to evaluate the detection accuracy of the graft ostia since, as pathological

structures, the number of graft ostia varies. We may have multiple detections around a true graft ostium, or multiple true graft ostia may be merged into one detection. We use the Hungarian algorithm to establish the one-to-one correspondence between the detection graft ostia and the ground truth [9]. A true ostium is claimed to be detected correctly, if its distance to the matched detection has a distance less than a threshold. Otherwise, it is mis-detected. If a detection has a distance larger than a threshold to the matched ground truth, it is claimed to be a false positive. Similar to native coronary ostium detection evaluation, we set the threshold to 5 mm since it is the approximate diameter of a coronary around its ostium.

Using the prior distribution on the anterior ascending aorta, we detect 351 graft ostia (95.6% detection rate) with only about 0.17 false positives per volume (row *Constrained Search on Aorta* of Table I). The false positives are mainly caused by surgical clips left on the aorta surface (see Fig. I). The mean distance of the correct detections to the ground truth is 1.61 mm. A few detection examples are shown in Fig. 4. For comparison, if we perform detection on the whole ascending aorta without using the prior distribution learned from the training set, we get 3.74 false positives/volume at the same detection rate (row *Search on Whole Aorta* of Table I). We also tried to detect graft ostia using the method presented in Section 2, which is effective to detect the anatomical native coronary ostia. As shown by the row *Search in Whole Volume* of Table I, the result is poor with a lower detection rate (81.4%) and a significant higher false positive rate (10.25 dections/volume).

Our approach compares favorably to the state-of-the-art. For example, Henemuth *et al.* [3] reported a detection rate of 57% for native coronary ostia and 27% for graft ostia on 61 datasets. The success rate of Wang and Smedby's method [4] for coronary ostia detection is only about 80% on 16 datasets.

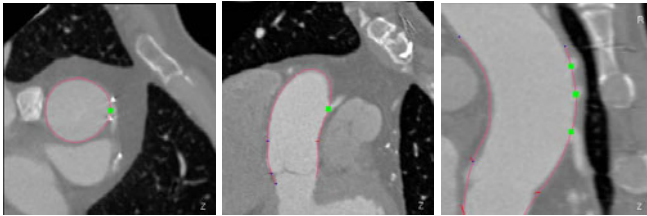


Fig. 4. Detection results for graft ostia of bypass coronaries on three datasets. The red contour shows the segmented ascending aorta and the green dots show the detected graft ostia. Note: The last dataset has three bypass coronaries.

Table 1. Comparison of graft ostium detection accuracy on 224 datasets with a total of 367 true graft ostia

	# Detections	Detection Rate	False Positives/Volume
Search in Whole Volume	307	81.4%	10.25
Search on Whole Aorta	352	95.9%	3.74
Constrained Search on Aorta	351	95.6%	0.17

On average, the proposed method takes only 0.40 seconds to process one volume on a computer with 2.33 GHz quad-core processors and 3 GB memory. It spends 0.09 s for image preprocessing, 0.10 s to detect the native coronary ostia, 0.09 s to segment the ascending aorta, and 0.12 s to detect graft coronary ostia on the aorta surface. Our approach is significantly faster than the previous methods, e.g., 6 s reported by [5] and 88.8 s reported by [4].

5 Conclusion

In this paper, we proposed an efficient method to automatically detect both native coronary ostia and graft ostia of bypass coronaries, using totally different treatments of the anatomical structures (native coronaries) and pathological structures (bypass coronaries). It is an important step forward to make the whole coronary analysis workflow fully automatic, which can increase the throughput of a physician and reduce the inter-user variability.

References

1. Lloyd-Jones, D., Adams, R., Carnethon, M., et al.: Heart disease and stroke statistics – 2009 update. *Circulation* 119(3), 21–181 (2009)
2. Metz, C., Schaap, M., van Walsum, T., van der Giessen, A., Weustink, A., Mollet, N., Krestin, G., Niessen, W.: 3D segmentation in the clinic: A grand challenge II — coronary artery tracking. In: *Proc. MICCAI Workshop on 3D Segmentation in the Clinic: A Grand Challenge* (2008)
3. Hennemuth, A., Boskamp, T., Fritz, D., Kühnel, C., Bock, S., Rinck, D., Scheuring, M., Peitgen, H.O.: One-click coronary tree segmentation in CT angiographic images. *Computer Assisted Radiology and Surgery*, 317–321 (2005)
4. Wang, C., Smedby, O.: An automatic seeding method for coronary artery segmentation and skeletonization in CTA. *The Insight Journal* (2008)
5. Tek, H., Gulsun, M.A., Laguitton, S., Grady, L., Lesage, D., Funka-Lea, G.: Automatic coronary tree modeling. *The Insight Journal* (2008)
6. Zheng, Y., Barbu, A., Georgescu, B., Scheuring, M., Comaniciu, D.: Four-chamber heart modeling and automatic segmentation for 3D cardiac CT volumes using marginal space learning and steerable features. *IEEE Trans. Medical Imaging* 27(11), 1668–1681 (2008)
7. 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)
8. Tu, Z.: Probabilistic boosting-tree: Learning discriminative methods for classification, recognition, and clustering. In: *Proc. Int’l. Conf. Computer Vision*, pp. 1589–1596 (2005)
9. Liu, G., Haralick, R.M.: Optimal matching problem in detection and recognition performance evaluation. *Pattern Recognition* 35(3), 2125–2139 (2002)

A *Holistic* Approach for the Detection of Media-Adventitia Border in IVUS

Francesco Ciompi^{1,2}, Oriol Pujol^{1,2}, Carlo Gatta^{1,2},
Xavier Carrillo³, Josepa Mauri³, and Petia Radeva^{1,2}

¹ Dep. of Applied Mathematics and Analysis, University of Barcelona, Spain

² Computer Vision Center, Campus UAB, Bellaterra, Barcelona, Spain

³ University Hospital “Germans Trias i Pujol”, Badalona, Spain
fciompi@maia.ub.es

Abstract. In this paper we present a methodology for the automatic detection of media-adventitia border (MAB) in Intravascular Ultrasound. A robust computation of the MAB is achieved through a holistic approach where the position of the MAB with respect to other tissues of the vessel is used. A learned quality measure assures that the resulting MAB is optimal with respect to all other tissues. The mean distance error computed through a set of 140 images is $0.2164 (\pm 0.1326)$ mm.

1 Introduction

Coronary arteries are mainly formed by three layers: *intima*, *media* and *adventitia*. Furthermore, two characteristic membranes can be defined, the Internal Elastic Lamina (IEL) and the External Elastic Lamina (EEL), separating the media from the adventitia layer. In healthy arteries, the shape of the vessel, delineated by the EEL, is approximatively circular, while in unhealthy cases it can become irregular, depending on the plaque amount and positioning. The clinical condition of the artery is commonly characterized by parameters like the *Plaque plus media Cross Sectional Area* (CSA), the *Maximum plaque plus media thickness* and the *Plaque Burden*. All of these measurements require the delineation of the media-adventitia border (MAB).

Intravascular Ultrasound (IVUS) is a catheter-based imaging technique commonly used during percutaneous interventions. It allows to visualize the internal morphology of the vessel, representing a suitable tool for the detection of the MAB. The media layer in IVUS appears as a dark ring surrounding the vessel (Fig. 1(a)). The thickness of the ring depends on vessel condition, varying from thick region ($800 \mu\text{m}$) in healthy cases to subtle line ($300 \mu\text{m}$) in unhealthy cases, due to the presence of plaque, that pushes the *media* layer towards the *adventitia*. Furthermore, the vessel border continuity is often hindered by the guide-wire artifact and by the shadow produced by calcifications. The weak appearance of the *media* makes the detection of the MAB by IVUS a difficult task.

Several approaches for the automatic MAB detection have been presented so far. In [1], a shape space, together with both calcification and bifurcation detection are used to determine the final border; in [2], anisotropic filtering

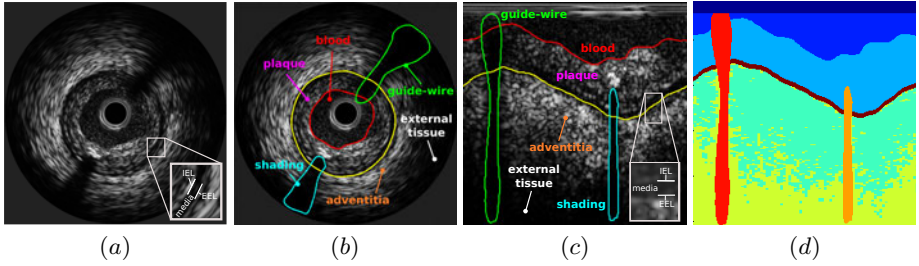


Fig. 1. Example of IVUS image in short-axis view (a); the manually labeled regions are depicted in both cartesian (b) and polar (c) view. In (d) the ground truth labels obtained by processing the borders in (c) are depicted.

operators and classification techniques are combined; in [3], an active contour model is used to regularize the information extracted from the contrast along the edges of the IVUS image; in [4], a Radial Basis Function is applied to the response of a low-pass filtering on the IVUS image; in [5], a snake using gradient vector flow regularizes the map of edges computed by the Canny operator; in [6], a Fast-Marching method based on gray level distributions, assumed as a mixture of Rayleigh pdfs is presented; in [7], the morphological characteristics of the IVUS image are used to initialize an active contour model; finally, in [8], deformable models are used, with energy function minimized by a neural network through a simulated annealing scheme.

In this paper we present a robust technique for the automatic detection of the MAb. The methodology is based on two assumptions: (1) only few regions in the image can be recognized, even by an expert physician, as belonging to the media; (2) the MAb can be robustly determined by exploiting the spatial distribution of the information on the morphology of the vessel. For these reasons, the proposed methodology is based on two steps. First, regions belonging to media layer are detected by means of a classification technique. Then, the MAb is detected by considering the relative position of the curve with respect to the surrounding tissues. To the best of our knowledge, this is the first method for the detection of MAb in IVUS based on the relationship between tissues. In particular, not only the local small neighborhood is considered, but all the tissues in the image. As a consequence, the detection of the MAb is the result of the contribution of several informations. This represents the *holistic* nature of the approach.

2 Method

Reflecting the two aforementioned assumptions, the proposed technique consists of two parts. First, the classification of the main morphological regions of the vessel in the IVUS image is performed by means of a multi-class Discriminative Random Fields (DRF) [9]. The usefulness of this step is two-fold. First, by defining the *media* as one of the classes, an initial approximation of the MAb is obtained (see Fig. 2(b)). Furthermore, the classification of the main regions of the vessel is a necessary information to guide the definition of the MAb where the

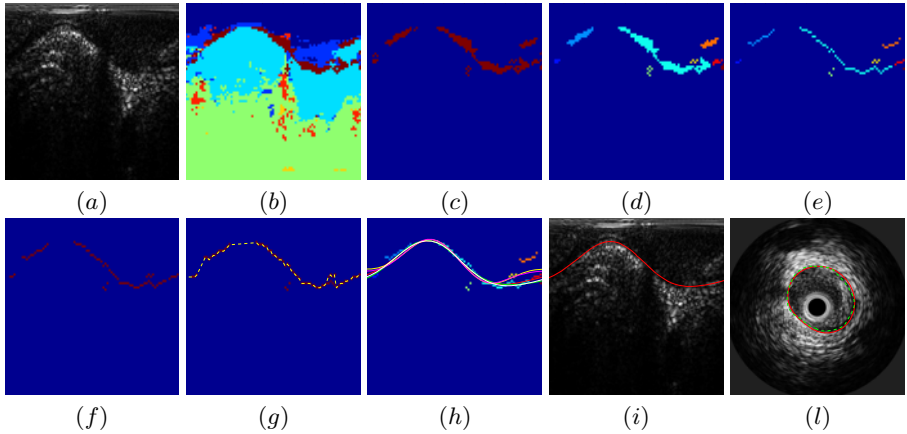


Fig. 2. Example of the main steps for the MAb detection in polar view (*a-i*). Each color in (b) represents a class; different colors in (d-e) represent independent isolated media regions. In (l) the detected border (continuous line) and the manual annotation (dotted line) are depicted in the short-axis view.

media appearance is weak or affected by artifacts. The regions labeled as *media* are often affected by two intrinsic errors: (1) due to the presence of artifacts, the media points are not continuous; (2) *media* regions can be detected inside the plaque region. Since the adventitia layer is mainly composed by fibrotic tissue, the DRF model learns that the media layer is placed between plaque and fibrosis. For this reason the luminal interface of fibrotic plaque may be labeled as *media*.

In cases where the MAb position is ambiguous, several guessings can be made. Among them, only one represents the solution to the problem, the optimal curve. Our hypothesis is that the optimal MAb is the one that exhibits the correct relative position with respect to all other tissues. In the second part of the method the points labeled as *media* are used as initial information with the aim of fitting a *continuous* and *smooth* curve. The *continuity* requirement is first fulfilled by interpolating the discontinuous media regions. Then, the *Fourier Series* is applied in order to obtain a *smooth* curve. The relationship between the curve and the tissues is expressed by a quality measure modeled on training examples: the optimal curve is the one producing the highest quality value. The proposed methodology is now explained in details.

2.1 Multi-class Tissue Classification

Tissues definition. A set of IVUS frames is manually labelled and used to train a multi-class classifier to discriminate the following classes: *blood*, *plaque*, *media*, *adventitia*, *guide-wire*, *shadowing* and *external tissue* (see Fig. 1 for details). The regions corresponding to these classes are derived by the expert manual annotations of *lumen border*, *External Elastic Lamina border* (EEL), *guide-wire* artifact and *shadowing* effect (due to calcifications). The regions are defined as follows (see Fig. 1(b)): **blood** (L) is the area inside the lumen border; **adventitia** (F)

is the external area limited by the EEL; **plaque** (P) is the region comprised between lumen and adventitia; **guide-wire** (G)/**shadow** (S) area enclosed in the guide/shadow border; **media** (M) is defined as a strip of constant thickness τ internal to the EEL border. In order to assure a sufficient amount of reliable *media* examples during the training process, the value of τ has been set as *half the thickness* of typical unhealthy media layers (around 150-200 μm). Finally, the class **external tissue** (B) consists of the set of points labeled as *adventitia* belonging to a cluster identified by low gray-level value and farther radial position. This procedure can be performed in a similar way in the polar view (see Fig. 1(c-d)).

Features. Both textural features and spectral features are extracted from IVUS data, since it has been demonstrated that the combination of both types of information provides an accurate tissue description [10]. To this aim, the raw radio frequency data (RF) of each frame is used to form the IVUS image, as detailed in [10]. As in [11], ten textual features are then extracted. Furthermore, the Fast Fourier Transform is applied to the RF data frame, resulting into 32 spectral features, as suggested in [12]. For each position of the image a feature vector $\mathbf{x} \in \mathbb{R}^{42}$ is obtained.

Classification. Given the presence of noise and artifacts, the tissues labeling in IVUS is a difficult task. In order to achieve a robust tissue classification, we propose to model the IVUS data as a Discriminative Random Fields (DRF) [9]. In this way the inference process can take into account local information about adjacent tissues. For the design of multi-class DRF, we adopt the recently proposed ECOC-DRF technique [13]. ECOC-DRF is a methodology for the design of both *node* and *edge* potential functions for DRF based on the Error-Correcting Output Codes (ECOC) framework [14]. As illustrated in [13], given the number of classes K for the node potential, the distance function for the edge potential is defined over a set of K^2 classes. In order to reduce the number of dichotomies for the edge potential, while maintaining the properties of the distance function in the ECOC space, we adopt the *Sublineal coding* technique using binary matrix and *Gray coding* technique, as suggested in [15]. By computing inference over the ECOC-DRF model, the Maximum A-Posteriori probability (MAP) over the seven classes is obtained, and used as input for the MAb detection (Fig. 2(b)).

2.2 Media-Adventitia Border Detection

In order to robustly detect the media layer we propose to exploit the information related to the *context* of the image; we want to take profit of the strong *a-priori* knowledge about the relative position of the tissues. To this aim, we designed a problem specific quality measure Φ , that takes into account the relative position of the curve with respect to surrounding tissues. The procedure to detect the MAb consists in (1) modeling a set of possible curves and (2) computing a quality measure for each curve. The MAb corresponds to the curve with the highest quality measure.

Firstly, each one of the M isolated regions labeled as media in the MAP (Fig. 2(c)) is considered as an independent element of a set (Fig. 2(d)). Among

them, only a certain subset delineates the correct vessel border: a region selection process is then required. For this purpose, a number of D (with $D \leq M$) regions are considered. Since we are interested in detecting the EEL, for each region, only the points belonging to the farther radial position are used (Fig. 2(e)). Then, the k -permutations of D (without repetitions) are computed, where $k = 1, \dots, D$, resulting in $\gamma = D + \sum_{k=2}^D \frac{D!}{2(D-k)!}$ combinations. Each combination consists in a set of regions delineating a profile that can be incomplete (Fig. 2(f)). Since the MAb is a continuous smooth curve, the detected border must fulfill continuity and smoothing constrains. In order to accomplish both requirements, the profile is first approximated by means of a piece-wise linear function (Fig. 2(g)). To assure a smooth curve, by exploiting the periodicity of the function, the Fourier series expansion is computed, and only the first $H \in \mathbb{N}^+$ harmonics are used. A set of γ smooth and continuous borders $b \in B$ is obtained (Fig. 2(h)).

For each curve $b \in B$ in polar coordinates, the normalized amounts of the tissues T positioned *above* (T_a) and *below* (T_b) with respect to the curve are computed. Furthermore, let us define $\mathbf{t}(b)^+$ and $\mathbf{t}(b)^-$ respectively the vector of correctly and wrongly placed tissues, with respect to the curve b . In order to define which are the significant tissues in the MAb detection problem, the following assumptions are done: (1) media points, already used in the border approximation, may be erroneously present in the plaque region; (2) the guide-wire artifact is observable in both region above and below the curve (see Fig. 1(d)). Hence, the classes *media* and *guide-wire* do not represent significant markers for curve positioning. For this reason, with reference to the notation in section 2.1, we define $\mathbf{t}(b)^+ = [L_a, P_a, S_b, B_b, F_b]$ and $\mathbf{t}(b)^- = [L_b, P_b, S_a, B_a, F_a]$. With these assumptions, a quality measure for the curve b is defined as follows:

$$\Phi(b) = \sum_{i=1}^m w_i^+ t_i(b)^+ + \sum_{j=1}^n -w_j^- t_j(b)^- = \mathbf{w}^T \cdot \mathbf{t}(b) \quad (1)$$

where $\mathbf{w} = [w_1^+, \dots, w_m^+, w_1^-, \dots, w_n^-]$ is a vector that weights the importance of different contributions to the quality measure. The solution is $\tilde{b} = \operatorname{argmax}(\Phi(b))$.

The estimation of the vector \mathbf{w} is a critical point; its role is two-fold: (i) to assign the proper contribution to each tissue, but also (ii) to take into account of wrong contributions due to errors in the labels. For this purpose, based on training examples, we compute $\mathbf{w}_{\text{opt}} = \operatorname{argmax} \sum_v \mathbf{w}^T \cdot \mathbf{t}^v(b_{GT})$, where $\mathbf{t}^v(b_{GT})$ is the vector of tissues of the v^{th} example in a validation dataset and b_{GT} is the *ground truth* border provided by experts. Given that the optimization problem is a maximization problem of an unbounded function, it is reasonable to constrain the values of the weights as $\|\mathbf{w}\|_2^2 = 1$, assuring non trivial solutions. The problem is solved by means of *Lagrange multipliers*.

3 Validation

Experimental Settings. A set of 18 in-vivo pullbacks have been acquired using a Galaxy II IVUS equipment (Boston Scientific) with a catheter Atlantis SR Pro 40 MHz (Boston Scientific). RF data have been sampled with a 12 bit sampling

Table 1. State-of-the-art methods vs holistic approach (HoliMAB). Details on the number of patients and images used in the validation are provided for all the methods, as well as the ultrasound frequency (MHz) of the used technology. MDE = mean distance error; MADE = maximum distance error; MAE = mean area error; IOV = inter observer variability.

	6	11	7	2	HoliMAB	IOV
Equipment	Jomed	Volcano	Clearview	Clearviews	Galaxy II	- -
# Patients	9	18	9	11	18	- -
# Images	540	647	50	5400	140	- -
MHz	20	20	40	40	40	- -
MDE (<i>mm</i>)	0.12 (0.07)	0.09 (0.16)	-	0.2265 (0.0688)	0.2164 (0.1326)	0.1226 (0.1654)
MADE (<i>mm</i>)	-	0.98 (1.78)	0.07 (0.78)	0.5715 (0.2296)	0.5162 (0.3209)	0.3717 (0.3117)
MAE (<i>mm</i> ²)	-0.2 (2.1)	0.75 (1.60)	0.17 (4.58)	8.6032 (3.3436)	0.5635 (0.4830)	0.3489 (0.4405)

card at a sampling rate of 200MHz. The most representatives and not consecutive challenging frames for each pullback have been selected and manually segmented by three experts, resulting in 140 frames.

Due to computational complexity in performing inference in a graphical model, we construct the lattice by reducing each feature map by a factor $\frac{1}{W_n^2}$. We assign the median value of each feature contained into non-overlapping blocks of $W_n \times W_n$ pixels of the image to each node of the graph. Thus, the parameter W_n represents a trade-off between computational cost and quality of the MAB estimation. As in [13] the edge feature function is set to $\tilde{\mathbf{x}} = |\mathbf{x}_i - \mathbf{x}_j|$. ECOC-DRF has been trained according to the Leave-One-Patient-Out (LOPO) scheme [10]: at each fold, one of the patients has been used for testing, and the rest is used for training. For the training process, the manual annotations of the expert exhibiting the minimal mean inter-observer variability are used. The used coding technique is *OneVsOne* for the node potential and *Sublineal* coding for the edge potential, requiring 21 and 12 dichotomies, respectively. The margin classifier for the ECOC-DRF has been set as suggested in [13]. The used *decoding* technique is the Attenuated Euclidean Distance for the node potential and the Euclidean Distance for the edge potential; the parameters (α_N, α_E) are set as described in [13]. The inference is performed by means of Belief Propagation (BP) [16].

Results Three error measures have been computed: (1) Mean Distance Error (MDE), computed by averaging the absolute difference between automatic border and manual annotation per A-line; (2) MAXimum Distance Error (MADE), computed by averaging the error on the maximum distance for each frame; (3) Mean Area Error (MAE), computed by averaging the absolute area differences. Parameters tuning, according to the LOPO scheme, provided $D = 8$ and $H = 2$. In order to determine D , the media regions have been ranked according to their length. As expected, the number of Fourier harmonics is small. Since the number of harmonics is inversely proportional to the regularization degree, this result confirms: (i) the hypothesis of assuming a continuous smooth curve and (ii) the shape of the vessel is often ellipsoidal. Figure 3 shows some graphical examples of border detection. Table 1 shows the performance of the proposed method when compared with other state of the art approaches and with the Inter Ob-

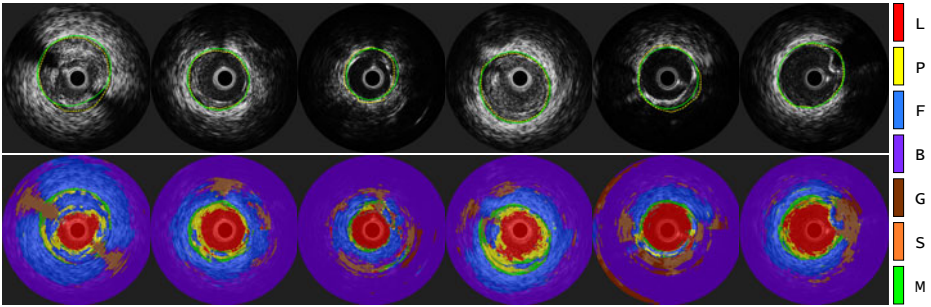


Fig. 3. Examples of automatic MAb detection (above) and corresponding classification maps (below). Continuous line represents the automatic segmentation; dotted line represents the manual annotation; labels names are assigned as in section 2.1.

server Variability. The performance of the *HoliMab* method is obtained using $W_n = 3$: this value experimentally demonstrates to be a reasonable trade-off between quality and computational cost.

4 Discussion

The comparison in Table 1 refers to the results provided in the cited paper: the database used in each case is different. In order to perform a fair comparison it is worth to note that some of the methods work on a 20 MHz US signal, while others work on 40 MHz. The difference between the two signals is remarkable, since the 20 MHz US exhibits higher tissue penetration power, while losing in local resolution. This fact increases the discrimination of tissues while losing details (but also artifacts) in IVUS data. The differences between the two technologies is evident since the errors obtained with 20 MHz techniques are smaller than IOV. To corroborate this assertion, it is worth to note that in [1] the authors claim that their approach does not work on a 40MHz technology.

The standard deviation of all the considered error measures in *HoliMab* is comparable with the inter-observer variability. This fact demonstrates the robustness of the proposed method that, exploiting the contextual information of the tissues positions, limits the range of error in the detection of the curve. The MDE value is comparable with the approach proposed in [2] that, given the experimental setup, can be considered as an exhaustive clinical study. The value of the MDE error can be controlled by two parameters: (1) the W_n value introduces a systematic error due to the change of scale; (2) the number of harmonics H controls the degree of regularization of the curve. The error introduced by the two parameters also represents a limitation of the method. In few cases where a calcification or a dense fibrosis surrounds almost completely the vessel, the method could perform sub-optimally. Finally, it is worth to note that by ranking the obtained values of Φ , several MAb can be proposed to physicians. Alternative solutions can be selected in case the first one is not satisfactory, avoiding the manual delineation of a better MAb.

5 Conclusions and Future Work

A novel methodology for the media-adventitia border detection based on context-analysis has been presented. The algorithm consists in a constant number of steps, tunable through the parameters (W_n, D) , that provide a trade-off between the detection quality and the execution time: for this reason, it can potentially run in real time. The method achieved results comparable with state-of-the-art and with inter-observer variability. The whole formulation of the media-adventitia border detection is based on RF data analysis: the method is then totally independent on the IVUS equipment technology and brand. Nevertheless, given the scarce accessibility to such data, the analysis of the usefulness of RF-based features through features selection is a straightforward step in our research. Furthermore, the applicability of the methodology to 20MHz IVUS data as well as exploiting spatial information in the vessel morphology by using consecutive frames is part of our future lines. It is worth to note that, though applied to the specific MAb detection problem, the functional in Eq. 11 is of general applicability for the segmentation of each tissue in medical imaging and, more in general, of any object in images where a frontier can be defined.

References

1. Unal, G., Bucher, S., Carlier, S., Slabaugh, G., Fang, T., Tanaka, K.: Shape-driven segmentation of the arterial wall in intravascular ultrasound images. *TITB* 12(3), 335–347 (2008)
2. Gil, D., Hernandez, A., Rodriguez, O., Mauri, J., Radeva, P.: Statistical strategy for anisotropic adventitia modelling in ivus. *TMI* 25, 768–778 (2006)
3. Zhu, H., Liang, Y., Friedman, M.H.: Ivus image segmentation based on contrast. In: *SPIE*, vol. 4684, pp. 1727–1733 (2002)
4. Papadogiorgaki, M., Mezaris, V., Chatzizisis, Y.S., Giannoglou, G.D., Kompatsiaris, I.: Automated ivus contour detection using intensity features and radial basis function approximation. In: *CBMS* (2007)
5. Sanz-Requena, R., Moratal, D., Garca-Snchez, D.R., Bod, V., Rieta, J.J., Sanchis, J.M.: Automatic segmentation and 3d reconstruction of intravascular ultrasound images for a fast preliminar evaluation of vessel pathologies. *CMIG* 31(2), 71–80 (2007)
6. Cardinal, M.H.R., Meunier, J., Soulez, G., Maurice, R.L., Therasse, E., Cloutier, G.: Intravascular ultrasound image segmentation: a three-dimensional fast-marching method based on gray level distributions. *TMI* 25(5), 590–601 (2006)
7. Giannoglou, G.D., Chatzizisis, Y.S., Koutkias, V., Kompatsiaris, I., Papadogiorgaki, M., Mezaris, V., Parissis, E., Diamantopoulos, P., Strintzis, M.G., Maglaveras, N., Parcharidis, G.E., Louridas, G.E.: A novel active contour model for fully automated segmentation of intravascular ultrasound images: In vivo validation in human coronary arteries. *CBM* 37, 1292–1302 (2007)
8. Plissiti, M.E., Fotiadis, D.I., Michalis, L.K., Bozios, G.E.: An automated method for lumen and media-adventitia border detection in a sequence of ivus frames. *TITB* 8(2), 131–141 (2004)
9. Kumar, S., Hebert, M.: Discriminative random fields. *IJCV* 68(2), 179–201 (2006)
10. Ciompi, F., Pujol, O., Gatta, C., Rodriguez-Leor, O., Mauri-Ferre, J., Radeva, P.: Fusing in-vitro and in-vivo intravascular ultrasound data for plaque characterization. *IJCI* 26, 763–779 (2010)

11. Ciompi, F., Pujol, O., Fernandez-Nofrerias, E., Mauri, J., Radeva, P.: Ecoc random fields for lumen segmentation in radial artery ivus sequences. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5762, pp. 869–876. Springer, Heidelberg (2009)
12. Sathyanarayana, S., Carlier, S., Wenguang, L., Thomas, L.: Characterization of atherosclerotic plaque by spectral similarity of radiofrequency intravascular ultrasound signals. *EuroIntervention* 5, 133–139 (2009)
13. Ciompi, F., Pujol, O., Radeva, P.: A meta-learning approach to conditional random fields using error-correcting output codes. In: IEEE International Conference on Pattern Recognition (ICPR), pp. 710–713 (2010)
14. Dietterich, T.G., Bakiri, G.: Solving multiclass learning problems via error-correcting output codes. *JAIR* 2, 263–286 (1995)
15. Bautista, M.A., Baro, X., Pujol, O., Radeva, P., Vitria, J., Escalera, S.: Compact evolutive design of error-correcting output codes. In: ECML (2010)
16. Yedidia, J.S., Freeman, W.T., Weiss, Y.: Understanding belief propagation and its generalizations. (TR-2001-22), pp. 239–269 (2002)

Orientation Histograms as Shape Priors for Left Ventricle Segmentation Using Graph Cuts

Dwarikanath Mahapatra and Ying Sun

Department of Electrical and Computer Engineering,
National University of Singapore, 4 Engineering Drive 3, Singapore 117576
dmahapatra@gmail.com, elesuny@nus.edu.sg

Abstract. Poor contrast in magnetic resonance images makes cardiac left ventricle (LV) segmentation a very challenging task. We propose a novel graph cut framework using shape priors for segmentation of the LV from dynamic cardiac perfusion images. The shape prior information is obtained from a single image clearly showing the LV. The shape penalty is assigned based on the orientation angles between a pixel and all edge points of the prior shape. We observe that the orientation angles have distinctly different distributions for points inside and outside the LV. To account for shape change due to deformations, pixels near the boundary of the prior shape are allowed to change their labels by appropriate formulation of the penalty and smoothness terms. Experimental results on real patient datasets show our method's superior performance compared to two similar methods.

1 Introduction

Segmentation of the left ventricle (LV) from dynamic contrast enhanced (DCE) magnetic resonance (MR) images of the heart is important for the analysis of cardiac functions. As an effective non-invasive imaging protocol, MRI has acquired great importance. Except for images with contrast enhancement, most images from the cardiac sequence show poor contrast between LV blood pool and myocardium wall. This, in addition to low spatial image resolution, makes segmentation of the LV a very challenging task when using only low level information (e.g. intensity, gradient, etc). This highlights the need for prior shape information in LV segmentation. In this paper we propose a graph cut based method to segment the LV blood pool and myocardium from DCE-MR image sequences using orientation histograms for prior shape information.

Previous works on segmentation using shape priors include, among others, active contours [1] and graph cuts [2,3]. Graph cuts have the advantage of being fast, give globally optimal results and are not sensitive to initialization [4], while active contours are sensitive to initialization and can get trapped in local minima. Shape information in graphs is based on interaction between graph nodes (or image pixels). Inter-pixel interaction is generally limited to the immediate neighborhood although graph cuts can handle more complex neighborhoods [4].

Therefore prior shape models assume great significance. We restrict further discussion to methods using prior shape information in graph cuts. The first works to use prior shape information in graph cuts were [23]. In [2] the zero level set function of a shape template of natural and medical images was used with the smoothness term to favor a segmentation close to the prior shape. Slabaugh et al. in [3] used an elliptical shape prior, under the assumption that many objects can be modeled as ellipses. However they employ many iterations where a pre-initialized binary mask is updated to get the final segmentation. Vu et al. [5] use a discrete version of shape distance functions to segment multiple objects, which can be cumbersome. A flux-maximization approach was used in [6], while in [7] the smoothness cost was modified to include star shape priors.

Other methods have focused on segmenting the LV using statistical shape information. Mitchell et al. in [8] use a multistage active appearance model to segment LV and right ventricle (RV). Besbes et al. [9] used a control point representation of the LV prior and other images were deformed to match the shape prior. Shape knowledge has also been combined with dynamic information to account for cardiac shape variability [10,11]. Ali et al. in [12] construct a shape prior with some degree of variability and use it to segment DCE-MR kidney images.

We propose a method that has the capability to handle different shapes and uses a single image to get the shape prior. The novelties of our work are two-fold. First, we determine the shape penalty based on the distribution of orientation angles between a pixel and edge points on the prior shape. The angle distribution makes the metric robust and invariant to the number of points on a shape. Second, our method uses a single image from each dataset to get prior shape information. In Section 2 we describe our method in greater detail. Section 3 presents results of our experiments on real datasets, and we list our conclusions in Section 4.

2 Theory

We use a second order Markov random field (MRF) energy function. MRFs are suitable for discrete labeling problems, and graph cuts can find their globally optimal solution for binary labels [4]. The energy function is given as

$$E(L) = \sum_{s \in P} D(L_s) + \lambda \sum_{(s,t) \in N} V(L_s, L_t), \quad (1)$$

where P denotes the set of pixels, L_s denotes label of pixel $s \in P$ and N is the set of neighboring pixel pairs. L_s denote the segmentation class of a pixel (0 for background and 1 for object). The labels of the entire set of pixels are denoted by L . $D(L_s)$, is a unary data penalty function derived from observed data and measures how well label L_s fits pixel s . V is a pairwise interaction potential that imposes smoothness and measures the cost of assigning labels L_s and L_t to neighboring pixels s and t . $\lambda = 0.01$ is a weight that determines the relative contribution of the two terms. Note that both D and V consist of two terms, each incorporating intensity and shape information.

Overview of method: Our method has the following steps: 1) choose a reference image for each dataset and identify a small region inside and outside the LV. These regions give the reference intensity histograms of object and background. 2) Segment the image using only intensity information. This gives an oversegmented LV which is used to update the intensity distributions of object and background. The initial segmentation need not be optimal; and 3) incorporate shape information in D and V to get the final segmentation

Intensity Information: Patches are identified on the LV and background and their intensity distributions are modeled as Gaussians by estimating the mean and variance for each label. The intensity penalty (D_I) is the negative log likelihood given by

$$D_I(L) = -\log Pr(I_s|L_s), \quad (2)$$

where I_s is the intensity at pixel s , Pr is the likelihood, and L_s =object/background is the label. The intensity smoothness term V_I assigns a low penalty at edge points based on the intensity of neighboring pixel pairs, and favours a piecewise constant segmentation result. It is defined as

$$V_I(L_s, L_t) = \begin{cases} e^{-\frac{(I_s-I_t)^2}{2\sigma^2}} \cdot \frac{1}{\|s-t\|}, & L_s \neq L_t, \\ 0 & L_s = L_t \end{cases} \quad (3)$$

σ determines the intensity difference up to which a region is considered as piecewise smooth. It is equal to the average intensity difference in a neighborhood w.r.t pixel s . $\|s-t\|$ is the Euclidean distance between s and t .

Shape Information: A prior shape is used to incorporate shape knowledge into $D(L_s)$. In our method the shape penalty is determined at each pixel using only a single image from each dataset. Figure 1(g) shows an illustration of the reference shape as a continuous circle and different points ($A-F$) inside and outside the shape. We shall explain the significance of the dotted line later. Figure 1(h) shows the reference image from a typical dataset with the LV endocardium outlined in yellow and the epicardium outlined in green. First we explain how the shape prior is used to segment the LV blood pool. Later we extend our explanation to the segmentation of both blood pool and myocardium. Say for point A (Fig. 1(g)), we calculate the orientation angles of the line joining it to every point on the outline of the reference shape and plot the distribution of these angles. We notice that points inside and outside the reference shape have different distribution characteristics. The $atan2$ function is used which returns angle values in the range $[-180, 180]$.

All points outside the shape (i.e., points $A-B$ and $D-F$) have the orientation angles distributed over two or three quadrants of the angle space. The line joining A to the closest point in the reference shape is the zero degree axis. For example, in the case of points outside the shape, the orientation angles are distributed over quadrants I, IV , or II, III (see Figs. 1(a)-(b), (d)-(f)) while for point C (which is inside the shape) the orientations are distributed over quadrants I, II, III, IV (Fig. 1(c)). For histograms the horizontal axis shows the angle range from -180°

to $+180^\circ$. For some points which are very close to the LV endocardium boundary but outside it, the distribution may be spread over three quadrants (I, II, IV or I, III, IV) but never all the four quadrants. This discriminating feature acts as a basis for assigning penalty values to the pixel, i.e., if the angle distribution is spread over two or three quadrants then it is likely to belong to the background and therefore

$$\begin{aligned} D_S(L_s = 0) &= 0, \\ D_S(L_s = 1) &= k_1. \end{aligned} \quad (4)$$

D_S is the shape penalty; $k_1 > 1$ is a constant of high value so pixel s does not take that label. If the distribution of angles is spread over four quadrants then the shape penalty is defined as

$$\begin{aligned} D_S(L_s = 0) &= k_1, \\ D_S(L_s = 1) &= 0. \end{aligned} \quad (5)$$

The above formulations makes the segmented shape very similar to the reference shape. Since the heart is made of muscular tissue there is bound to be elastic deformations in them with a resulting change in shape. In order to segment the deformed LV we relax the constraints on pixels near the prior's boundary. Referring back to Fig. 1 we observe that pixel A is nearer to the reference shape than point B . We set a threshold normal distance of d_{th} pixels from the shape within which the LV may deform and allow for a change in labels of pixels within this d_{th} distance. Pixels which fall within this area are equally likely to take a particular label, and therefore have the same penalty. Therefore,

$$D_S(L_s = 0/1) = k_2, \quad d_s \leq d_{th}, \quad (6)$$

Here $0 < k_2 < k_1$, indicates that both labels are equally likely. d_s refers to the distance of pixel s from the shape. To determine d_{th} we adopt the following steps. After rigid alignment we choose 4 datasets (240 images) with large deformations (expecting it to cover a large range of shape deformations) and manually segment the endocardium region in all frames. For each dataset we compute the mean shape and distance of each shape from the mean. It is observed that the maximum distance of any individual shape from the average shape is 5 mm and the average distance is 3.5 mm. Therefore we set $d_{th} = 5mm$. Note that for any point lying on the edge of the prior shape $d_s = 0$. For most such edge points the distribution of orientation angles may not be spread over four quadrants. But, since $d_s = 0$, the penalty is assigned according to Eqn. (5).

Another contribution of our work is the formulation of the smoothness penalty based on the prior shape. Although the extent of interaction is usually limited to the immediate neighborhood, an arbitrary neighborhood can also be used [4]. For any pixel within d_{th} distance, it has some finite interaction with the closest pixel on the prior shape. This interaction is more important than the data penalty value to determine which pixels near the prior's boundary may change labels (due to deformations) as pixels within d_{th} distance of the contour have equal penalty for both labels. If such a pixel is very similar to the prior-shape's edge

pixels (based on some feature), it is likely to have the same segmentation label. Thus we define V_S as

$$V_S(L_s, L_t) = \begin{cases} 1 - h, & L_s \neq L_t, \\ 0 & L_s = L_t. \end{cases} \quad (7)$$

where $0 < h \leq 1$ is the difference of orientation angle histograms given by the χ^2 metric. The total energy function is thus defined as

$$E(L) = \sum_{s \in P} [w_1 D_I(L_s) + D_S(L_s)] + \lambda \sum_{(s,t) \in N} [V_I(L_s, L_t) + V_S(L_s, L_t)], \quad (8)$$

$w_1 = 0.6$ decides the relative contribution of intensity information to the penalty and was set empirically.

Orientation histograms tend to overcome the shortest path problem inherent to graph cuts [7]. Therefore they are used in calculating V_S . To save further computation time the penalty values are calculated based on these orientation histograms without extra distance transform calculations. A single point in a bin is considered for spread in a quadrant. V_S uses pixels within d_{th} distance of the prior while V_I uses only neighboring pixels.

Extension to Multiple Classes: Segmentation of the image into LV blood pool, myocardium and background requires 3 labels. Those points which lie outside the continuous circle in Fig. 1 (g) but inside the dotted circle correspond to the myocardium. All points outside the dotted circle (points G and H) are on

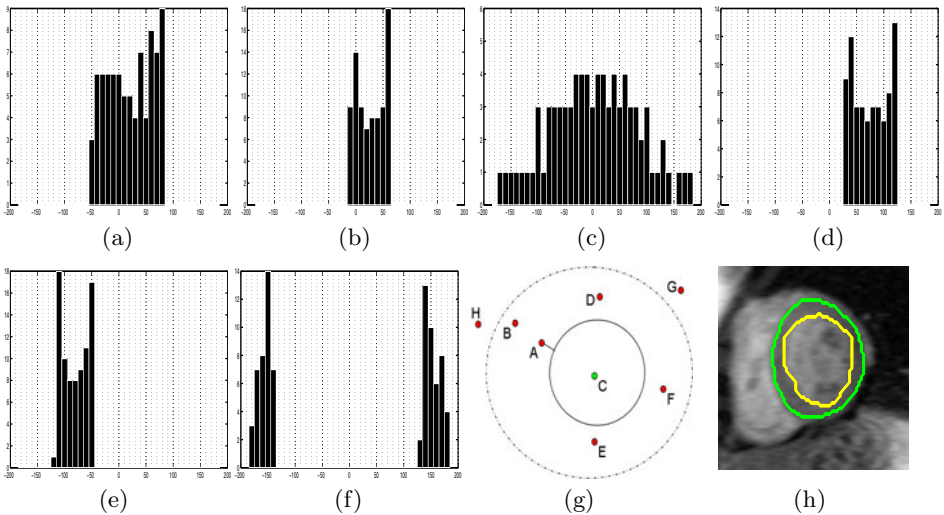


Fig. 1. Illustration of shape prior segmentation using orientation information. Orientation histograms for (a) point A ; (b) point B ; (c) point C ; (d) point D ; (e) point E ; (f) point F ; (g) synthetic image showing the different points outside and inside the shape; (h) reference cardiac image with LV endocardium highlighted in yellow and epicardium in green;

the background. Depending upon the location of a pixel, different penalty values for each label is assigned. The value of d_{th} is the same as before.

3 Experiments and Results

Cardiac images were acquired on Siemens Sonata MR scanners following bolus injection of Gd-DTPA contrast agent. The pixel spacing ranges from $(1.5 \times 1.5) - (2.8 \times 2.8) \text{mm}^2$. The acquired datasets were all in 2D and a total of 7 datasets were used to test our method. Each dataset had 60 frames with a total of 420 images. The images were rigidly registered to the prior shape before segmentation. We have 3 segmentation labels for LV blood pool, LV myocardium and background. The LV blood pool and myocardium were manually segmented in all frames by experts. Automatic segmentations were obtained using four methods: graph cuts with intensity information alone (*GC*); our method using shape priors with graph cuts (*GCSP*); the methods in [26]. The automatic segmentations were compared with manual segmentation using Dice Metric (*DM*) [13]. In all our experiments $k_1 = 10$, $k_2 = 1.2$ and $w_1 = 0.6$.

The method in [2] was implemented using the zero level set of the shape template in the smoothness cost for prior shape knowledge. For implementing [6] we calculate the appearance and location prior as described in their work, while for the shape prior we use the flux maximization and template based star shape constraint only. We weight the different components of the energy function to get the best segmentation results.

Figure 2 shows frames from different stages of contrast enhancement in a typical cardiac perfusion sequence alongwith the segmentation results from different methods. The outline of segmented masks are shown in different colours, and the manual segmentation is shown in red. These images exhibit poor contrast between the LV and the surrounding myocardium such that relying on intensity information alone leads to over-segmentation of the LV. The regions of inaccurate segmentation by each method is highlighted by yellow arrows. A visual inspection of the results indicates improved segmentation accuracy when using shape priors. Further, it also indicates that our method is more robust to shape changes. The formulation of our penalty and smoothness terms allows near-boundary pixels the flexibility to change their labels. If the reference shape is close to the average shape of the sequence, the accuracy does not change much. However if it has large deformation compared to other frames, lower *DM* values are obtained.

Table 1 shows the quantitative segmentation performance for all four methods. Use of shape priors significantly improves the segmentation accuracy. Intensity information alone is not sufficient for LV segmentation in MR images. Our method shows highest *DM* values and lowest contour distances, indicating that orientation histograms can very effectively account for shape changes due to deformations. Use of shape prior information in both penalty and smoothness term improves our method's accuracy over others. The average time taken for segmentation by different methods is 13 s (*GC*), 29 s (*GCSP*), 17 s [2] and 32 s

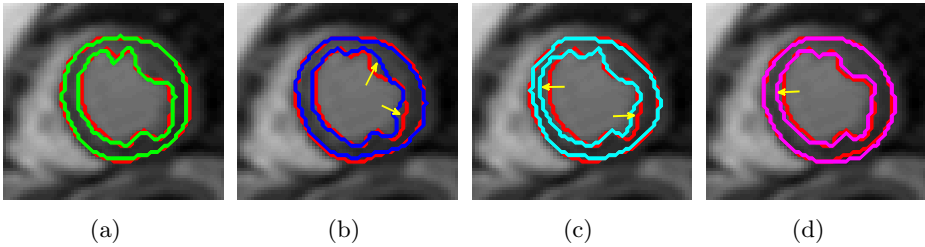


Fig. 2. Segmentation results using different methods. The manual segmentation is shown in red in all images and the results for automatic segmentation is shown in different colors. Results for (a) *GCSP*; (b) *GC*; (c) [2]; and (d) [6]. Areas of inaccurate segmentation are shown by yellow arrows.

Table 1. Comparative performance of segmentation accuracy using four methods and two metrics (*DM* and contour distance). The values show the average measures over all datasets. *GC*-graph cut with intensity only; *GCSP* is our method using graph cuts and shape priors; methods in [2], [6]. Values indicate the mean and standard deviation.

	Dice Metric (%)				Mean Absolute Distance (mm)			
	<i>GC</i>	<i>GCSP</i>	[2]	[6]	<i>GC</i>	<i>GCSP</i>	[2]	[6]
Epicardium	87.1±1.2	95.1±0.6	92.8±1.3	93.1±0.9	3.2±0.	0.7±0.3	1.4±0.4	1.2±0.2
Endocardium	88.2±0.9	94.7±1.1	93.1±0.5	92.4±1.0	2.8±0.5	0.8±0.4	1.2±0.3	1.3±0.4
Overall LV	87.7±1.1	95.0±0.9	92.9±0.8	92.±0.9	3.0±0.6	0.7±0.3	1.3±0.3	1.3±0.3

for [6]. The time is for images of dimension $75 - 82 \times 83 - 87$ using MATLAB 7.5 on a PC with Pentium 4, 3 GHz dual core processor. We did not optimize the code to take advantage of the multi core processor.

Influence of parameters k_1 , k_2 , w_1 . w_1 decides the relative contribution of intensity to the total penalty. If $w_1 > 0.9$ then D_I 's contribution is high compared to D_S and the final labels are influenced solely by D_I . On the other hand if $w_1 < 0.3$ D_S dominates and D_I has no influence on the final labels. $w_1 = 0.6$ is empirically chosen to give the best results for a wide range of datasets. k_1 should have value sufficiently higher than zero to impose hard constraints for labeling. k_2 should be low enough to not dominate over V_S , and also be greater than zero to ensure unambiguous solutions. If $k_2 > 1.5$ the contribution of V_S towards the final labels decreases.

4 Conclusion

In this paper we have proposed a novel shape prior segmentation method using graph cuts for segmenting the LV from a sequence of dynamic cardiac perfusion images. Manual intervention is limited to identifying the LV in a reference image to get the prior shape. The shape penalty is calculated using the distribution of orientation angles from every pixel to the edge points of the prior shape.

Penalty and smoothness terms are formulated such that pixels near the boundary of the shape prior can change their labels to account for shape change due to deformations. When combined with the intensity distributions of the object and background, our method results in accurate segmentation of the LV for low contrast perfusion MR images. Experimental results on real patient datasets show the advantages of using shape priors, and the superior performance of our method over two related methods. In future work we aim to extend the method for more segmentation classes and other imaging modalities.

References

1. Cremers, D., Tischhauser, F., Weickert, J., Schnorr, C.: Diffusion snakes: Introducing statistical shape knowledge into the Mumford-Shah functional. *Intl. J. Comp. Vis.* 50(3), 295–313 (2002)
2. Freedman, D., Zhang, T.: Interactive graph cut based segmentation with shape priors. In: *IEEE CVPR*, pp. 755–762 (2005)
3. Slabaugh, G., Unal, G.: Graph cuts segmentation using an elliptical shape prior. In: *IEEE ICIP*, pp. 1222–1225 (2005)
4. Boykov, Y., Veksler, O.: Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell.* 23, 1222–1239 (2001)
5. Vu, N., Manjunath, B.: Shape prior segmentation of multiple objects with graph cuts. In: *IEEE CVPR*, pp. 1–8 (2008)
6. Chittajallu, D., Shah, S., Kakadiaris, I.: A shape driven mrf model for the segmentation of organs in medical images. In: *IEEE CVPR*, pp. 3233–3240 (2010)
7. Veksler, O.: Star shape prior for graph cut segmentation. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) *ECCV 2008, Part III*. LNCS, vol. 5304, pp. 454–467. Springer, Heidelberg (2008)
8. Mitchell, S., Lelieveldt, B., van der Geest, R., Bosch, H., Reiver, J., Sonka, M.: Multistage hybrid active appearance models: Segmentation of cardiac mr and ultrasound images. *IEEE Trans. Med. Imag* 20(5), 415–423 (2001)
9. Besbes, A., Komodakis, N., Paragios, N.: Graph-based knowledge-driven discrete segmentation of the left ventricle. In: *IEEE ISBI*, pp. 49–52 (2009)
10. Zhu, Y., Papademetris, X., Sinusas, A., Duncan, J.: Segmentation of left ventricle from 3d cardiac mr image sequence using a subject specific dynamic model. In: *Proc. IEEE CVPR*, pp. 1–8 (2008)
11. Sun, W., Çetin, M., Chan, R., Reddy, V., Holmvang, G., Chandar, V., Willsky, A.: Segmenting and tracking of the left ventricle by learning the dynamics in cardiac images. In: Christensen, G.E., Sonka, M. (eds.) *IPMI 2005*. LNCS, vol. 3565, pp. 553–565. Springer, Heidelberg (2005)
12. Ali, A., Farag, A., El-Baz, A.: Graph cuts framework for kidney segmentation with prior shape constraints. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) *MICCAI 2007, Part I*. LNCS, vol. 4791, pp. 384–392. Springer, Heidelberg (2007)
13. Pluempitiwiriwajew, C., Moura, J., Wu, Y., Ho, C.: STACS: new active contour scheme for cardiac MR image segmentation. *IEEE Trans. Med. Imag.* 24(5), 593–603 (2005)

Myocardial Segmentation of Late Gadolinium Enhanced MR Images by Propagation of Contours from Cine MR Images

Dong Wei¹, Ying Sun¹, Ping Chai², Adrian Low², and Sim Heng Ong¹

¹ Department of Electrical and Computer Engineering,
National University of Singapore, Singapore
{dongwei, elesuny, eleongsh}@nus.edu.sg

² Cardiac Department, National University Heart Centre, Singapore

Abstract. Automatic segmentation of myocardium in Late Gadolinium Enhanced (LGE) Cardiac MR (CMR) images is often difficult due to the intensity heterogeneity resulting from accumulation of contrast agent in infarcted areas. In this paper, we propose an automatic segmentation framework that fully utilizes shared information between corresponding cine and LGE images of a same patient. Given myocardial contours in cine CMR images, the proposed framework achieves accurate segmentation of LGE CMR images in a coarse-to-fine manner. Affine registration is first performed between the corresponding cine and LGE image pair, followed by nonrigid registration, and finally local deformation of myocardial contours driven by forces derived from local features of the LGE image. Experimental results on real patient data with expert outlined ground truth show that the proposed framework can generate accurate and reliable results for myocardial segmentation of LGE CMR images.

1 Introduction

Viability assessment of the myocardium after the experience of a myocardial infarction is essential for diagnosis and therapy planning. In particular, the detection, localization and quantification of infarcted myocardial tissue, also called infarct/infarction/scar, are important for determining whether and which part(s) of a heart that has undergone a myocardial infarction may benefit from re-vascularization therapy. In a typical CMR examination, a contrast agent is injected and a cine sequence is acquired approximately at the same time; 15 to 20 minutes later, an LGE scan is performed, and by then scars exhibit brighter (enhanced) intensities than healthy myocardium. This is because scars accumulate more contrast agent and experience delayed wash-in and wash-out of contrast agent. Typical LGE images with enhanced scars can be seen in Fig. 4.

Delineation of myocardial contours is the first step in locating and quantifying infarctions in LGE images. Because manual delineation is not only time-consuming but also subject to intra- and inter-observer variability, it is highly desirable to automate the contour delineation process. However, automatic segmentation of myocardial contours is a difficult task, due mainly to the intensity

inhomogeneity of the myocardium resulting from the accumulation of contrast agent in infarcted areas. To the best of our knowledge, there has been little research aimed at fully automatic myocardial segmentation in LGE images, and there is no commercially or publicly available automatic segmentation tool for clinical use. Most of the existing approaches utilize pre-delineated myocardial contours in the corresponding cine MRI as *a priori* knowledge [1,3]. Such an approach is reasonable because the patient is asked to stay still during the entire acquisition process and there are many methods available for automatic segmentation of cine MRI [5]. Nevertheless, major difficulties in this approach include: (i) misalignment and nonrigid deformation between cine and LGE data due to respiratory motion and/or imperfectness of electrocardiography gating (which will be worse in cases of arrhythmia); and (ii) differences in resolution, field of view, and global intensity of cine and LGE data.

In this paper, we propose a multi-level framework for automatic myocardial segmentation of LGE CMR images in a coarse-to-fine manner. Our work is different from the aforementioned ones [1,3] in three respects: (i) we fully utilize shared information between corresponding cine and LGE images by registering the cine image to the LGE image in an affine-to-nonrigid manner, including both shape and intensity information; (ii) instead of using conventional similarity metrics such as mutual information and cross-correlation, experimentally we choose pattern intensity [9] which leads to accurate nonrigid registration of corresponding cine and LGE images; and (iii) at the finest level of segmentation, we propose to detect endocardial edges by adaptively selecting one of the two cases: normal endocardium and sub-endocardial scars, as well as, incorporate a new effective thickness constraint into the evolution scheme based on the simplex mesh [2] geometry. The rest of the paper is organized as follows. Section 2 describes the proposed automatic segmentation framework. Section 3 presents the experimental results on real patient data, followed by the conclusion in Section 4.

2 Method

Our automatic segmentation framework comprises three major steps from coarse to fine levels: (i) estimate an affine transformation by maximizing normalized cross-correlation (NCC) for rough alignment of corresponding cine and LGE images; (ii) nonrigidly register the affine-transformed cine image to the LGE image using B-spline based free-form deformation (FFD) [7] with pattern intensity [9] as the similarity metric; and (iii) further deform the myocardial contours obtained from the previous step based on local features of the LGE image using the simplex mesh geometry [2]. Figure 1 shows the segmentation results obtained at the different stages of our framework.

2.1 Pre-processing and Affine Transformation

In this work, we only consider short-axis slices. A cine sequence comprises multiple slices at different phases, covering a complete cardiac cycle. Unlike cine data,

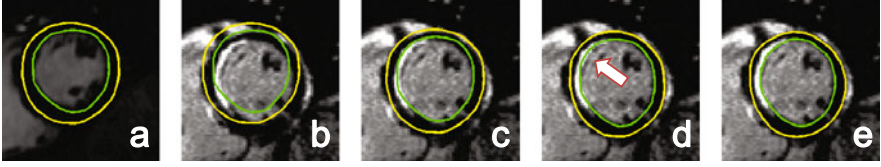


Fig. 1. The segmentation results obtained at the different stages of our framework: (a) the *a priori* segmentation in the cine image; (b) direct overlay of the *a priori* onto the LGE image without any processing; (c) after affine transformation; (d) after b-spline based FFD; (e) after local deformation of contours. Note that from (d) to (e) a small defect in the *a priori* is corrected (as highlighted by the arrow).

a set of LGE data comprises slices at only one phase, which is often located between end-systole and end-diastole. Therefore, given a target LGE image, the first task is to select a corresponding cine image. We select the cine image with the same spatial location and the closest phase to the LGE image according to their respective DICOM header information, and delineate myocardial contours in the selected cine image as *a priori* segmentation. Then we normalize each pair of corresponding cine and LGE images to: (i) the same physical resolution; (ii) similar histogram distributions (specifically, we specify the histogram of the LGE image to resemble that of the cine image); and (iii) the same size.

The coarse level segmentation is achieved by registering the cine image to the target LGE image through a constrained affine transformation. For the 2D case the affine transformation matrix in homogeneous coordinates can be expressed as

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ 0 & 0 & 1 \end{bmatrix}. \quad (1)$$

Based on the assumption that there should not be any significant rotation or shearing effects, and the scaling and translation effects should also be small, we constrain the estimated affine transformation so that:

$$|a_{11} - 1|, |a_{22} - 1| < \varepsilon_{\text{scale}} \quad \text{and} \quad |a_{13}|, |a_{23}| < \varepsilon_{\text{translate}}, \quad (2)$$

where $\varepsilon_{\text{scale}}$ and $\varepsilon_{\text{translate}}$ are corresponding thresholds and set to 0.1 and 10 for our data. Here we use NCC as the similarity measure because it is invariant to both shift and scale intensity distortions [4] and hence can overcome systematic intensity variations between corresponding cine and LGE images. For a fast implementation, we adopt the enhanced cross-correlation (ECC) algorithm proposed in [4], which optimizes an efficient approximation of NCC and leads to a closed-form solution in each iteration.

2.2 B-Spline Based Nonrigid Registration

Since the deformation of a heart cannot be completely described by an affine transformation, we apply B-spline based nonrigid registration [7] following the

affine transformation. For a 2D image, B-spline based FFD is controlled by a mesh of control points $\Phi = \{\phi_{i,j}\}$. Let $s_x \times s_y$ denote the uniform spacing of the mesh grid, then the FFD can be written as:

$$\mathbf{T}(x, y, \Phi) = \sum_{l=0}^3 \sum_{m=0}^3 B_l(u)B_m(v)\phi_{i+l,j+m}, \quad (3)$$

where $i = \lfloor x/s_x \rfloor - 1$, $j = \lfloor y/s_y \rfloor - 1$, $u = x/s_x - \lfloor x/s_x \rfloor$, $v = y/s_y - \lfloor y/s_y \rfloor$, and B_l is the l th basis function of B-splines. The transformation field \mathbf{T} completely determines deformation of the whole image. Because B-splines are locally controlled, they are relatively more efficient compared to other kinds of splines such as thin-plate splines, even for a large number of control points.

Different from most registration methods which use conventional similarity measures such as normalized mutual information (NMI), experimentally we select *pattern intensity* [9] as the similarity measure for the nonrigid registration. Given two images I_1 and I_2 , pattern intensity operates on the difference image $I_{\text{diff}} = I_1 - I_2$. If I_1 and I_2 are two images of the same object and well registered, structures from this object should vanish and there should be a minimum number of structures or patterns in I_{diff} . A suitable similarity measure should, therefore, characterize the *structuredness* of I_{diff} . Pattern intensity considers a pixel of I_{diff} to belong to a structure if it has a significantly different value from its neighboring pixels. Using a constant radius r to define the neighborhood of the pixel being examined, pattern intensity is defined as:

$$\begin{aligned} P_{r,\sigma}(I_1, I_2) &= P_{r,\sigma}(I_{\text{diff}}) \\ &= \frac{1}{N_{I_{\text{diff}}}} \sum_{x,y} \frac{1}{N_r} \sum_{(x-v)^2+(y-w)^2 \leq r^2} \frac{\sigma^2}{\sigma^2 + [I_{\text{diff}}(x, y) - I_{\text{diff}}(v, w)]^2}, \end{aligned} \quad (4)$$

where $N_{I_{\text{diff}}}$ and N_r denote the number of pixels in I_{diff} and the neighborhood respectively. A constant σ is introduced to suppress the impact of noise.

Pattern intensity has the following desirable properties: (i) small deviations in intensity would still produce measurements remaining near to the maximum value due to the introduction of σ ; (ii) large differences of intensity values have the same effect regardless of their magnitude due to the asymptotic nature, making the measure robust to a few large differences in pixel intensity; and (iii) its regional nature is able to reduce the effect of differences with scales larger than r . These properties make it robust to the presence of enhanced infarctions of relatively small area in LGE images. Indeed, we have experimented with popular similarity metrics including squared intensity difference, NMI, and NCC, and found that pattern intensity gives best results for most LGE images in our database (see Fig. 2 for a demonstration). This is consistent with the findings from a comparative study of six similarity measures for intensity-based 2D-3D vertebra image registration [6], which state that pattern intensity outperformed the others and was able to achieve accurate registration even when soft-tissue structures and interventional instruments were present as differences between the images.

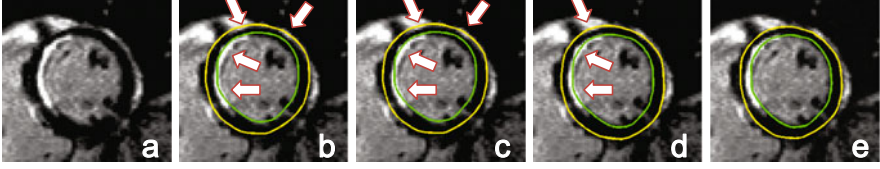


Fig. 2. A comparison of the effects of different similarity metrics on b-spline based cine and LGE images registration: (a) the LGE image; (b) squared intensity difference; (c) NCC; (d) NMI; (e) pattern intensity. The bright arrows in (b)-(d) highlight the locations where pattern intensity based registration outperforms others.

In our implementation, we use the optimization scheme described in [7]. The cost functional to be minimized consists of two terms, one for similarity and the other for smoothness. The similarity cost is defined as:

$$\mathcal{C}_{\text{similar}}[I_1, \mathbf{T}(I_2, \Phi)] = 1 - P_{r,\sigma}[I_1, \mathbf{T}(I_2, \Phi)], \quad (5)$$

and the smoothness cost $\mathcal{C}_{\text{smooth}}$ is defined as the 2-D bending energy of a thin-plate of metal to regularize \mathbf{T} [7]. The final cost functional is a weighted sum of $\mathcal{C}_{\text{similar}}$ and $\mathcal{C}_{\text{smooth}}$ with a weighting factor λ which controls the trade-off between accurate match of the two images and smoothness of the transformation:

$$\mathcal{C}(\Phi) = \mathcal{C}_{\text{similar}}[I_1, \mathbf{T}(I_2, \Phi)] + \lambda \mathcal{C}_{\text{smooth}}(\mathbf{T}). \quad (6)$$

We have experimentally determined the values for grid spacing s_x , s_y and weighting factor λ . We found that setting s_x , $s_y = 8$ and $\lambda = 0.2$ would produce relatively reasonable yet smooth segmentation for most LGE images in our database.

2.3 Local Deformation of Myocardial Contours

After the first two steps, we can already obtain reasonable segmentation results which are quite close to true myocardial boundaries. However, some minor flaws still exist, e.g. a fraction of the infarct is excluded from the myocardial contours (Fig. 1(d)). Thus we further locally deform the existing myocardial contours to ensure that scars are enclosed by the final contours, and improve segmentation accuracy. Different from the previous two steps in which segmentation is achieved by registration of corresponding cine and LGE images, in this step we directly deform the contours based on local features of the LGE image alone.

We use the *simplex mesh* [2] geometry and a Newtonian mechanical model to represent and deform the contours. We represent both endo- and epicardial contours with 80 mesh vertices. At each vertex we define three different forces which move it jointly, namely, smoothness $\mathbf{F}_{\text{smooth}}$, edge attraction \mathbf{F}_{edge} and myocardium thickness $\mathbf{F}_{\text{thick}}$. $\mathbf{F}_{\text{smooth}}$ imposes uniformity of vertex distribution and continuity of simplex angles [2]. \mathbf{F}_{edge} draws vertices towards detected myocardial edge points along radial directions of the left ventricle (LV). We detect the edge points by searching pixels along the radial directions in a limited range

(Fig. 3 (a)) and picking those with maximum intensity changes with respect to their neighbors on the search line. Specially, while searching for endocardial edge points, we consider two cases: edges of sub-endocardial scars (Fig. 3 (b)) and normal endocardium (Fig. 3 (c)). The former case is made much easier to detect nowadays thanks to the improvements in MRI acquisition techniques (which improve the contrast between the blood pool and sub-endocardial scars), and given a higher priority over the latter considering the prevalence of sub-endocardial scars in infarction patients. In addition, the search range is confined to a narrow band of 4 pixels, due to the closeness of the contours obtained in the previous step to real myocardial borders, and to avoid an unstable large extent of deformation. Letting \hat{p}_e denote the detected edge point for any vertex p , \mathbf{F}_{edge} can be expressed as:

$$\mathbf{F}_{\text{edge}} = \omega_p(\hat{p}_e - p), \quad (7)$$

where ω_p is a weight linearly and directly proportional to the local intensity change at \hat{p}_e and normalized to the interval $[0, 1]$. Finally, $\mathbf{F}_{\text{thick}}$ establishes a connection between nearest endo- and epicardial vertices along the radial directions (Fig. 3(d)), and aims to overcome the problem that arises when the edge search fails (e.g. when there is poor contrast between scars and the blood pool). Letting p_{endo}^0 denote any vertex on the endocardial contour before deformation, and p_{epi}^0 its nearest neighbor on the epicardial contour, we introduce $\mathbf{F}_{\text{thick}}$ as:

$$\begin{aligned} \mathbf{F}_{\text{epi,thick}} &= p_{\text{endo}}^t + (p_{\text{epi}}^0 - p_{\text{endo}}^0) - p_{\text{epi}}^t, \\ \mathbf{F}_{\text{endo,thick}} &= p_{\text{epi}}^t - (p_{\text{epi}}^0 - p_{\text{endo}}^0) - p_{\text{endo}}^t, \end{aligned} \quad (8)$$

where p_*^t is the vertex after t th iteration.

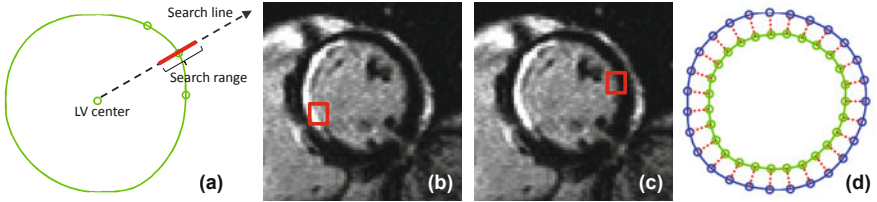


Fig. 3. An illustration of \mathbf{F}_{edge} and $\mathbf{F}_{\text{thick}}$: (a) the search for myocardial edges along radial directions of the LV; (b) an edge of sub-endocardial infarction highlighted with the red square; (c) an edge of normal endocardium highlighted with the red square; (d) connections established between endo- and epicardial contours as thickness constraint.

Letting p^t denote the position of any vertex at time t , the deformation scheme can be written as

$$p^{t+1} = p^t + (1 - \gamma)(p^t - p^{t-1}) + \alpha \mathbf{F}_{\text{smooth}} + \beta \mathbf{F}_{\text{edge}} + \theta \mathbf{F}_{\text{thick}}, \quad (9)$$

where γ is a damping factor. In each iteration, we first fix the endocardial contour and update the epicardial contour, and then update the endocardial contour with the epicardial contour fixed. Various weights in (9) are experimentally determined as: $\gamma = 0.7$, $\alpha = 0.35$, $\beta = 0.15$, $\theta = 0.1$.

3 Experimental Results

We have tested the proposed automatic segmentation framework on 59 short-axis LGE CMR images (with corresponding pre-segmented cine CMR images as *a priori* knowledge) from 10 patients that are clinically diagnosed as having experienced myocardial infarction. The images were acquired with a Siemens Symphony MRI scanner. Based on visual examination, we find that segmentation results produced by our framework are consistently correct and accurate for nearly all of the LGE images. Figure 4 shows some example results, together with manual segmentation by an expert. We have also conducted a quantitative evaluation of our framework by calculating distance errors and the Dice coefficient between the automatic results and the expert’s manual segmentation. The average and maximum errors between the automatic and manual contours are: 0.97 ± 0.45 and 2.60 ± 1.15 pixels for the endocardium, 0.89 ± 0.40 and 2.38 ± 0.98 pixels for the epicardium, and 0.93 ± 0.42 and 2.49 ± 1.06 pixels for both. This level of error is quite close to intra- and inter-observer variability. As to the area similarity, the Dice coefficient is $93.59 \pm 3.80\%$ for the endocardium, $95.63 \pm 2.62\%$ for the epicardium, and $82.49 \pm 9.27\%$ for the myocardium. Since the ultimate goal of automatic myocardial segmentation in LGE CMR images is to serve subsequent automatic quantification of scars, we have further implemented the automatic scar segmentation algorithm proposed in [8] to compare scars within the automatic and manual contours. The volumetric Dice coefficient between segmented scars within respective contours is $79.78 \pm 9.72\%$, which is comparable to those reported in [8].

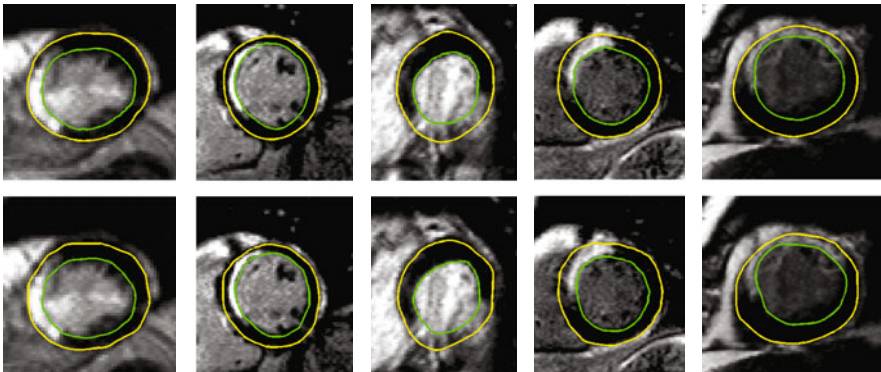


Fig. 4. Some example segmentation results of our automatic framework (top row), as compared with those by an expert (bottom row)

4 Conclusion

In this paper we present an automatic segmentation framework for LGE CMR data that fully utilizes shared information between corresponding cine and LGE images. Given myocardial contours in cine CMR images, our framework first

generates intermediate segmentation results by registering the cine image to the target LGE image in an affine-to-nonrigid manner, and then obtains final segmentation results by locally deforming the myocardial contours with forces derived from local features of the LGE image alone. For the nonrigid registration, different from most other registration methods, we experimentally chose pattern intensity as the similarity metric; it works robustly for most of the LGE images whereas popular similarity metrics such as NMI often cannot compare. For the local deformation of the myocardial contours, we use the simplex mesh geometry to represent the contours and displace each vertex according to the combination of three different forces: smoothness, edge attraction, and thickness. Our experimental results on real patient data, along with the quantitative evaluation, show that the proposed automatic segmentation framework can successfully segment short-axis LGE images with good accuracy and reliability.

References

1. Ciofalo, C., Fradkin, M., Mory, B., Hautvast, G., Breeuwer, M.: Automatic myocardium segmentation in late-enhancement MRI. In: IEEE ISBI 2008, pp. 225–228 (May 2008)
2. Delingette, H.: General object reconstruction based on simplex meshes. *International Journal of Computer Vision* 32, 111–146 (1999)
3. Dikici, E., ODonnell, T., Setser, R., White, R.D.: Quantification of Delayed Enhancement MR Images. In: Barillot, C., Haynor, D.R., Hellier, P. (eds.) MICCAI 2004. LNCS, vol. 3216, pp. 250–257. Springer, Heidelberg (2004)
4. Evangelidis, G., Psarakis, E.: Parametric image alignment using enhanced correlation coefficient maximization. *IEEE TPAMI* 30(10), 1858–1865 (2008)
5. Hautvast, G., Lobregt, S., Breeuwer, M., Gerritsen, F.: Automatic contour propagation in cine cardiac magnetic resonance images. *IEEE T-MI* 25(11), 1472–1482 (2006)
6. Penney, G., Weese, J., Little, J., Desmedt, P., Hill, D., Hawkes, D.: A comparison of similarity measures for use in 2-D-3-D medical image registration. *IEEE T-MI* 17(4), 586–595 (1998)
7. Rueckert, D., Sonoda, L., Hayes, C., Hill, D., Leach, M., Hawkes, D.: Nonrigid registration using free-form deformations: application to breast MR images. *IEEE T-MI* 18(8), 712–721 (1999)
8. Tao, Q., Milles, J., Zeppenfeld, K., Lamb, H., Bax, J., Reiber, J., van der Geest, R.: Automated segmentation of myocardial scar in late enhancement MRI using combined intensity and spatial information. *Magnetic Resonance in Medicine* 64(2), 586–594 (2010)
9. Weese, J., Buzug, T., Lorenz, C., Fassnacht, C.: An approach to 2D/3D registration of a vertebra in 2D x-ray fluoroscopies with 3D CT images. In: Troccaz, J., Grimson, E., Mösges, R. (eds.) CVRMED-MRCAS 1997. LNCS, vol. 1205, pp. 119–128. Springer, Heidelberg (1997)

Minimum Average-Cost Path for Real Time 3D Coronary Artery Segmentation of CT Images

Ning Zhu and Albert C.S. Chung

Lo Kwee-Seong Medical Image Analysis Laboratory,
Department of Computer Science and Engineering,
Hong Kong University of Science and Technology, Hong Kong
{nzhu, achung}@cse.ust.hk

Abstract. In this paper, we propose a Minimum Average-cost Path (MACP) model for segmenting 3D coronary arteries by minimizing the average edge cost along path in discrete 4D graph constructed by image voxels and associated radii. Prim's Minimum Spanning Tree method is used for efficient optimization of the MACP model. The centerline and the radii of the cross sections of the coronary artery are extracted simultaneously during the optimization. The method does not need any image preprocessing steps and has been intensively validated as an effective approach with the Rotterdam Coronary Artery Algorithm Evaluation Framework [1]. The computational cost of the proposed method is particularly low (7.467 seconds per segment, 18.5mm/s on average), which makes real time segmentation of coronary artery possible. Shortcut problem, which is a classic issue of the minimal path techniques, can also be overcome by the proposed method.

1 Introduction

In recent years, coronary artery segmentation has become an important research issue since it can help diagnosis and treatment of coronary diseases, which is one of the major causes of death around the world. As described in [4], a variety of methods have been proposed for vessel segmentation with the aim of extracting centerlines in 3D space by using classical minimal path algorithms. Some exceptions can be found in [5,7,11] and they share the same idea of incorporating the cross sectional radius as an additional dimension together with the coordinates of the voxels. Thus when the tracking procedure ends, the radii of the vessel is also obtained. The resultant paths are better centered by using these methods.

Similar to [5,7,11], our approach represents the coronary artery centerline with its radii as a 4D tree branch with 3D coordinates of the centerline point and one addition dimension for associated radii. Given the start point, the end point and a point inside the distal part of an artery, our method can trace the centerline with associated radii efficiently without taking the risk of tracing a shortcut path, which is a classical issue of minimal path techniques. Two kinds of shortcut problem are shown in Fig. 1. In extreme case, as shown in Fig. 1(a), the path may be traced outside the vessels (dashed line), while in other shortcut

cases (Fig. 1(b)), the path (dashed line) obtained by minimal path techniques may stay in the vessels but are less satisfactory than the real centerline path (solid line). The model proposed in this paper, namely Minimum Average-cost Path (MACP), concentrates on low average edge cost by minimizing the newly designed average cost of all edges on the path in the form of probability.

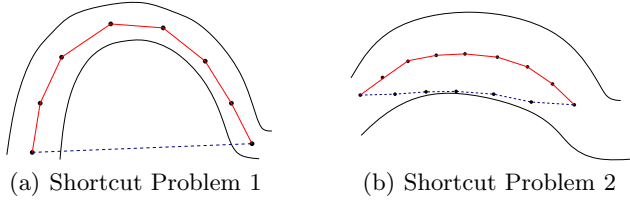


Fig. 1. Shortcut Problems of Minimal Path Approaches

To the best of our knowledge, there is only one work [9] on real time 3D CT coronary artery segmentation, but it achieves relatively low accuracy. The MACP model proposed in our paper introduces a new energy term for fast propagation and the proposed method tends to consider much less candidate voxels than minimal path methods. These properties of MACP make it possible to trace the coronary artery with a very low computational cost.

There are four favorable properties of the proposed method. First, according to the experimental results listed in the Rotterdam Coronary Artery Algorithm Evaluation Framework [1], our method is computationally more efficient than all other methods for segmenting 3D CT images of coronary artery. The second merit is that the classical issue of minimal path techniques, shortcut problem, can be overcome by the proposed method. Third, the proposed method does not need any preprocessing steps such as enhancement, and has been validated intensively in the experiments.

2 Minimum Average-Cost Path Model

Similar to [5,7,11], the cross sections of coronary arteries are assumed to be circular as an approximation which is reasonable for thin vessels such as coronaries. The 4D graph node $X_t = (\mathbf{p}_t, r_t)$ (Fig. 2), in the assumption includes the voxel position $\mathbf{p}_t = (x_t, y_t, z_t)$ and the radii r_t of the cross section. We let \mathbf{d}_t be the direction from the previous node position $\mathbf{p}_{t-1} = (x_{t-1}, y_{t-1}, z_{t-1})$ to the node position $\mathbf{p}_t = (x_t, y_t, z_t)$ being considered. The cross section is defined to be perpendicular to \mathbf{d}_t . Lesage *et. al.* proposed a Bayesian minimal path method

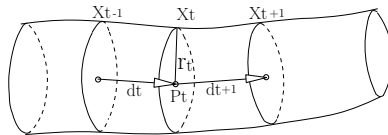


Fig. 2. Geometric Model

for coronary artery segmentation in [5]. The recursive update of the a posterior probability of a model realization is given by the Bayes' rule:

$$P(X_{[0:t]}|Y_{[0:t]}) \propto P(X_{[0:t-1]}|Y_{[0:t-1]})P(X_t|X_{t-1})\frac{P_v(Y_t)}{P_{bg}(Y_t)}, \tag{1}$$

where Y_t is the observation of X_t , $P_v(Y_t)$ and $P_{bg}(Y_t)$ are the probabilities of the response as a path node and background node respectively. $P(X_{[0:t]}|Y_{[0:t]})$ denotes the joint distribution of sequence of nodes on the path given the observations associated with the nodes. Maximizing Equation (1) is equivalent to minimizing its negative logarithm:

$$C(X_{[0:t]}) = C(X_{[0:t-1]}) - \log(P(X_t|X_{t-1})) - \log(P_v(Y_t)) + \log(P_{bg}(Y_t)) + M, \tag{2}$$

where $C(X_{[0:t]}) = -\log P(X_{[0:t]} | Y_{[0:t]})$, $X_t = (\mathbf{p}_t, r_t)$ and M is a constant ensuring that $C(X_{[0:t]})$ is positive.

As described in Section 1, the minimal path method in [5] may trace a shortcut path. The shortcut problem is caused by the assumption that the path can be obtained by maximizing $P(X_{[0:t]}|Y_{[0:t]})$, which is equivalent to minimizing $C(X_{[0:t]})$. This may lead to the problem of shortcut since the value of $C(X_{[0:t]})$ may still become smaller if fewer points are included into the path.

In order to ensure a low average cost path, the value of $P(X_t|Y_t, X_{[0:t-1]})$ should be large for each X_t in the route. We can assume $P(X_t|Y_t, X_{[0:t-1]}) = P(X_t|Y_t, X_{t-1})$ by regarding the 4D path as first order Markov chain. Maximizing $P(X_t|Y_t, X_{[0:t-1]})$ is equal to minimizing $-\log P(X_t|Y_t, X_{t-1})$. As such, we define a new notation ACP in our model:

$$\begin{aligned} \text{ACP}(\mathbf{v}) &= -\frac{\sum_{j=1}^t \log\{P(X_j | Y_j, X_{j-1})\}}{t} \\ &= -\frac{\sum_{j=1}^t \{\log\{P(X_{[0:j]} | Y_{[0:j]})\} - \log\{P(X_{[0:j-1]} | Y_{[0:j-1]})\}\}}{t}. \end{aligned}$$

\mathbf{v} here denotes the whole path from the start point X_0 to the end point X_t . Considering a single path \mathbf{v} , our model aims at minimizing $\text{ACP}(\mathbf{v})$. Suppose $C(X_{[0:j]}) = -\log(P(X_{[0:j]} | Y_{[0:j]})) = C(X_{[0:j-1]}) + \text{Edge}(j)$,

$$\text{ACP}(\mathbf{v}) = \frac{\sum_{j=1}^t \{C(X_{[0:j]}) - C(X_{[0:j-1]})\}}{t} = \frac{\sum_{j=1}^t \text{Edge}(j)}{t} = \frac{C(X_{[0:t]}) - C(X_0)}{t}.$$

According to Equation (2), we can obtain

$$\text{Edge}(j) = -\log(P(X_j | X_{j-1})) - \log(P_v(Y_j)) + \log(P_{bg}(Y_j)) + M.$$

In these steps above, we translate our intuition to the model MACP.

3 Bayesian Minimum Spanning Tree Implementation

3.1 A Novel Energy Term for Fast Propagation

In this section, an energy term $E(\mathbf{v})$ is defined for fast propagation and optimization of the MACP model. \mathbf{v} here denotes the whole path from the start

point S to the end point E .

$$E(\mathbf{v}) = L(\mathbf{v})/N + ACP(\mathbf{v}),$$

where N is the total number of nodes on the path except S , and

$$\begin{aligned} L(\mathbf{v}) &= -2 \times \text{Dis}(S, E) \\ &= \sum_{i=1}^N \{((\text{Dis}(p_{i-1}, S) - \text{Dis}(p_i, S)) + (\text{Dis}(p_i, E) - \text{Dis}(p_{i-1}, E)))\}, \end{aligned}$$

in which $p_0 = S$ and $p_N = E$, and $\text{Dis}(\cdot, \cdot)$ denotes the distance between two points. Assume

$$\text{LES}(i) = \text{LS}(i) + \text{LE}(i) = \text{Dis}(p_{i-1}, S) - \text{Dis}(p_i, S) + \text{Dis}(p_i, E) - \text{Dis}(p_{i-1}, E),$$

we can obtain,

$$E(\mathbf{v}) = \sum_{i=1}^N \text{LES}(i) + \text{Edge}(i)/N. \tag{3}$$

Since $L(\mathbf{v})$ is a negative constant, minimizing $L(\mathbf{v})/N$ is equal to minimizing N . By minimizing $E(\mathbf{v})$, ACP is minimized with the constrain of reducing the total number of points on the path. In each propagation phase, we choose the point with the minimum average edge cost from the start point to it according to Equation (4). \mathbf{v}_j here denotes the path from the start point to the point being considered and the number of points on the path is j :

$$E(\mathbf{v}_j) = \sum_{i=1}^j \{\text{LES}(i) + \text{Edge}(i)\}/j. \tag{4}$$

3.2 Implementation Detail

We modify the Prim’s algorithm to fit the proposed model. The overall algorithm is described in Algorithm 1.

In the experiments, one artery is traced in two parts given the start point S_p , the end point E_p and a point inside the distal part of the artery, A_p . First, we set $S = A_p, E = S_p$. With Algorithm 1, the path from A_p to S_p is obtained. Then we set $S = A_p, E = E_p$ to obtain the path from E_p to A_p . After that, the whole artery can be obtained by simply connecting the two parts together. The reason for not just tracing back from E_p to S_p is that the shape of the distal parts, especially the thin ones, are irregular. After step 13 in Algorithm 1, a tree is obtained. And then a low average cost path can be obtained by tracing back from E to S .

The proposed method tends to consider much less candidate voxels than the minimal path methods. For example, in Fig. 3, the path from node A to node T is the real path. Different with minimal path methods, the proposed method tends to dismiss the nodes such as $F_i (i = 1, \dots, n)$ since the average cost from A to F_i is large, and does not have to bother the paths to useless nodes $D_i (i = 1, \dots, n)$ and $E_i (i = 1, \dots, n)$. For minimal path method, the number of the useless nodes kept in the memory will increase exponentially while propagating towards the end point

Algorithm 1. Tracking Algorithm for MACP

- 1: **INPUT:**
 S – Start node position for tracking, E – End node position for tracking
 - 2: **INITIALIZATION:**
 Find the best radii r_s for S , Put $X_s = (S, r_s)$ into tree T , set $C_p = S$.
 - 3: **while** $C_p \neq E$ **do**
 - 4: **for** each neighboring node position N_j of C_p and N_j is not in T **do**
 - 5: **for** each different radii r_j related to node position N_j **do**
 - 6: **if** $E(v_j)$ is smaller than the largest $E(v_i)$ of record $X_i = (N_i, r_i)$ in QUEUE **then**
 - 7: Replace the record $X_i = (N_i, r_i)$ in QUEUE with $X_j = (N_j, r_j)$
 - 8: **end if**
 - 9: Update best candidate node outside QUEUE
 - 10: **end for**
 - 11: **end for**
 - 12: Let $X_p = (N_p, r_p) = \arg \min_{X_p \in \text{QUEUE}} E(v_p)$, $C_p = N_p$, put $X_p = (N_p, r_p)$ into T , replace $X_p = (N_p, r_p)$ with best candidate node outside QUEUE.
 - 13: **end while**
 - 14: Trace back from E to S in T to obtain the whole node path P .
 - 15: **OUTPUT:** P
-

T . So the proposed method can trace the vessel with a lower computational cost than minimal path methods and realize the real time segmentation.

During the propagation, the energy term $E(v_j)$ in Algorithm 1 is calculated by Equation (4). The prior term $P(X_t | X_{t-1})$ in constrained radius variations is described in [5]. Observations Y_t are the responses of a multi-scale oriented medialness feature and details can be found in [6].

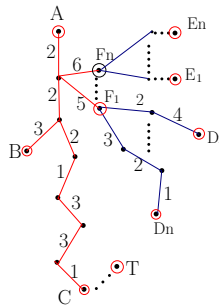


Fig. 3. Efficiency of MACP Model

4 Experiment and Validation

The centerlines of 24 testing datasets (96 vessels) were submitted to the Rotterdam Coronary Artery Algorithm Evaluation Framework [1] and the evaluation results are reported in Table 1. The overlap percentage between the experiment

Table 1. Results from the Rotterdam Framework

Measure	% / mm			score		
	min.	max.	avg.	min.	max.	avg.
OV	60.8%	100.0%	93.4%	34.3	100.0	66.6
OT	60.7%	100.0%	93.6%	30.4	100.0	67.7
AI	0.24 mm	0.70 mm	0.41 mm	17.5	54.2	28.3

results and the ground truth is denoted as OV, which is the most important measure in the framework. OT gives an indication of how well the method is able to track the clinically relevant vessel section with a diameter of 1.5 mm or larger. AI denotes the accuracy of centerline extraction, provided that the evaluated centerline is inside the vessel. The score is defined between 0 and 100, with 50 for a result of the order of the inter-observer variability. We note that, S_p , E_p and A_p are provided by the Evaluation Framework.

The relatively good performance of our method was confirmed with 93.4% overlap percentage (OV) on average, and the score for overlap percentage is 66.6, which means the proposed method exceeds inter-observer overlap with its score 50. The radii of the typical coronaries range from 0.5mm to 4.0mm, and the spacing of the images is about 0.3mm to 0.4mm in each direction, so the average result achieved (0.41mm) for measure AI is fine. To improve the AI measure, we can perform interpolation to the image and obtain more node positions. However, it will be slow to trace the path in the interpolated image. Thus it is a trade-off between AI and computational efficiency. OF denotes how much a coronary artery has been extracted before making an error, but it is meaningless for the proposed method and we omit it here. In the proposed method, the tracking starts from point A_p back to point S_p and from point A_p to E_p , so the measure OF actually cannot denote how much of a vessel has been extracted before making the first error.

To evaluate the computation time of the proposed method, we recorded the tracking time of the training datasets (32 vessels) with different Q values, Q is the size of QUEUE in Algorithm 1. As we can see in Table 2, the computation time increases monotonously with Q . For $Q = 20$, with which we obtain the 93.4% on OV, the average computation time is 7.467 seconds. According to [8], the reference standard centerlines are about 138mm on average, so the speed of tracing is 18.5mm/s. The processor used in the experiment is dual core AMD Opteron 2216 (2.4GHz) and the physical memory is 8GB. For running the program, the maximum memory used is 3GB even though the raw image and the gradient vector field for the whole image are restored in the memory.

Table 2. Computation Time (in seconds) of MACP

Size	min.	max.	avg.
Q=10	1.943	11.242	5.804
Q=20	2.015	16.137	7.467
Q=30	2.058	18.945	7.953
Average	2.005	15.441	7.075

Given centerlines and related radii, surfaces can be obtained. In Fig. 4(a), surface was rendered from reference centerlines and radii from [1]. Surface for the proposed method is shown in Fig. 4(b). Since the centerlines obtained by the proposed method are based on voxels, the number of voxels on the centerlines obtained is much less than the reference. Also, radius r_t associated with \mathbf{p}_t ranges from $r_{t-1} - 0.6\text{mm}$ to $r_{t-1} + 0.6\text{mm}$ with a fixed step of 0.3mm . These all may make it difficult to have a smooth rendering surface. However, the rendering surface for our method is satisfactory.

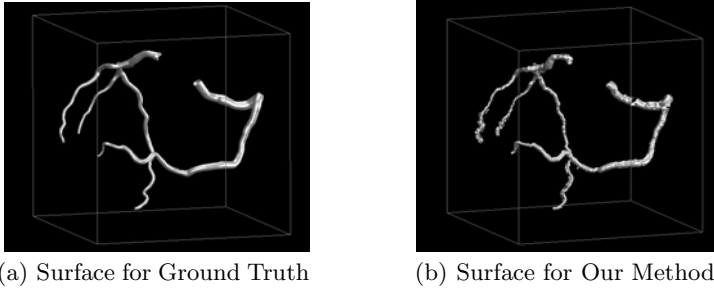


Fig. 4. Rendering Surface Results

Table 3 presents the experimental results of 15 state-of-the-art approaches in the Rotterdam evaluation framework [1]. As we can see in Table 3, the computational efficiency (7.467 seconds per segment, 29.868 seconds per dataset) of our method is better than all approaches. The only one approach [9], with its computation time comparable with the proposed method, however achieves a lower overlap performance OV(84.7%).

There are only four [5][3][10][2] out of fifteen approaches having better overlapping performance than the proposed method (93.4%). However, the proposed method is much faster than all of them. Meanwhile, these methods have other problems. In Friman's MHT method [3], more human intervene is required to see whether the tracking is finished. In [10], vessel enhancement is performed

Table 3. Experimental Results From [1]

Measure	OV	Computation Time
MHT(Friman et al., 2008) [3]	98.5%	6 minutes
BayesianMaxPaths (Lesage, 2008) [5]	97.5%	4 minutes
VirtualContrast2b(Wang et al., 2008) [10]	96.7%	2 minutes
Tracer (Szymczak, 2008) [2]	95.1%	30 minutes
GVFTubenLinkage (Bauer and Bischof, 2008)	92.7%	10 minutes
TwoPointMinCost(Metz et al., 2008)	91.9%	12 minutes
AxialSymmetry(Dikici et al., 2008)	90.8%	5 minutes
3DInteractiveTrack(Zhang et al., 2008)	89.6%	3-6 minutes
KnowledgeBasedMinPath (Krissian et al., 2008)	88.0%	7 hours
AutoCoronaryTree (Tek et al., 2008) [9]	84.7%	<30 seconds
DepthFirstModelFit (Zambal et al., 2008)	84.7%	4-8 minutes
CocomoBeach (Kitslaar et al., 2008)	78.8%	70 seconds
ElasticModel(Hernndez Hoyos et al.)	77.0%	2-6 minutes
VirtualContrast(Zhang et al.)	75.6%	2 minutes
CoronaryTreeMorphoRec (Castro et al., 2008)	67.0%	30 minutes

before tracking. Actually, it is impossible for the methods like [10] to run in real time if enhancement process is required since it is a time consuming process. The computational cost of [2] is much higher than the proposed method, however, the superiority of OV of the method is not so obvious (95.1% vs 93.4%). In contrast with the proposed method, [3,10,2] do not estimate the vessel radii and this leads to difficulty in rendering the vessel surfaces as we can have in Fig. 4(b). Although the overlap performance (OV) of methods [5,3] is high, the use of minimal path methods may lead to shortcut problem in some cases as shown in Fig. 1.

5 Conclusion

In this paper, we have proposed a MACP model for 3D coronary artery segmentation and presented the use of Prim's Minimum Spanning Tree method for efficient optimization of the MACP model. The centerline and the radii of the cross sections of the coronary artery are extracted simultaneously during the optimization on the discrete 4D graph. The method has been validated intensively with Rotterdam Coronary Artery Algorithm Evaluation Framework [1] and the tracking quality is good, and it does not need any image preprocessing steps. The computational cost is particularly low, and this makes real time segmentation of coronary artery possible. Shortcut problem, which is a classic issue of the minimal path techniques, can be overcome by the proposed method.

References

1. Rotterdam Coronary Artery Algorithm Evaluation Framework, <http://coronary.bigr.nl/>
2. Vessel tracking by connecting the dots, <http://hdl.handle.net/10380/1406>
3. Friman, O., Kühnel, C., Peitgen, H.: Coronary artery centerline extraction using multiple hypothesis tracking and minimal paths. *The MIDAS Journal* (2008)
4. Lesage, D., Angelini, E., Bloch, I., Funka-Lea, G.: A review of 3D vessel lumen segmentation techniques: Models, features and extraction schemes. *Medical Image Analysis* 13(6), 819–845 (2009)
5. Lesage, D., Angelini, E., Bloch, I., Funka-Lea, G.: Bayesian maximal paths for coronary artery segmentation from 3D CT angiograms. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) *MICCAI 2009*. LNCS, vol. 5761, pp. 222–229. Springer, Heidelberg (2009)
6. Lesage, D., Angelini, E., Bloch, I., Funka-Lea, G.: Design and study of flux-based features for 3D vascular tracking. In: *ISBI*. pp. 226–229 (2009)
7. Li, H., Yezzi, A.: Vessels as 4-D curves: Global minimal 4-D paths to extract 3-D tubular surfaces and centerlines. *IEEE TMI* 26(9), 1213 (2007)
8. Schaap, M., Metz, C., Walsum, T.v., et al.: Standardized evaluation methodology and reference database for evaluating coronary artery centerline extraction algorithms. *Medical Image Analysis* 13(5), 701–714 (2009)

9. Tek, H., Gulsun, M., Laguitton, S., Grady, L., Lesage, D., Funka-Lea, G.: Automatic coronary tree modeling. *The MIDAS Journal* (2008)
10. Wang, C., Smedby, Ö.: Integrating automatic and interactive methods for coronary artery segmentation: let the pacs workstation think ahead. *International Journal of Computer Assisted Radiology and Surgery*, 1–11
11. Wink, O., Niessen, W., Viergever, M.: Multiscale vessel tracking. *IEEE TMI* 23(1), 130–133 (2004)

Robust Surface-Based Multi-template Automated Algorithm to Segment Healthy and Pathological Hippocampi

Hosung Kim¹, Tommaso Mansi², Neda Bernasconi¹, and Andrea Bernasconi¹

¹ Department of Neurology and McConnell Brain Imaging Center, Montreal Neurological Institute and Hospital, McGill University, Montreal, Quebec, Canada

² Siemens Corporate Research, Image Analytics and Informatics, Princeton, NJ, USA
{khs001, neda, andrea}@bic.mni.mcgill.ca,
{tommaso.mansi}@siemens.com

Abstract. The most frequent drug-resistant epilepsy is temporal lobe epilepsy (TLE) related to hippocampal atrophy. In addition, TLE is associated with atypical hippocampal morphologies. Automatic hippocampal segmentations have generally provided unsatisfactory results in this condition. We propose a novel segmentation method (SurfMulti) to statistically estimate locoregional texture and shape using a surface-based approach that guarantees shape-inherent point-wise correspondences. To account for inter-subject variability, including shape variants, we used a multi-template library derived from a large database of controls and patients. SurfMulti outperformed state-of-the-art volume-based single- and multi-template approaches, with performances comparable to controls (Dice index: 86.1 vs. 87.5%). Furthermore, the sensitivity of SurfMulti to detect atrophy was similar to that of manual volumetry. Given that the presence of hippocampal atrophy in TLE predicts a favorable seizure outcome after surgery, the proposed automated algorithm assures to be a robust surrogate tool in the presurgical evaluation for the time-demanding manual procedure.

Keywords: multi-template, segmentation, surface-based texture, SPHARM-PDM, template library, hippocampus, MRI, epilepsy.

1 Introduction

The most frequent form of drug-resistant epilepsy is temporal lobe epilepsy (TLE) related to hippocampal sclerosis, which generally appears as atrophy on MRI. Detecting hippocampal atrophy is clinically relevant as it allows lateralizing the side of seizure focus and defining the surgical target. The removal of this pathology is the only effective treatment, offering seizure freedom in 80% of cases [1].

Manual labeling is considered the gold standard to measure hippocampal volume, as it is accurate, reproducible, and sensitive. High time requirement, rater-bias, and the demand to study large cohorts of healthy and diseased populations, however, have motivated the automation of hippocampal volumetry.

Most segmentation methods employ deformable model- [2], appearance- [3], or atlas-based approaches [4-6]. The use of deformable parametric surfaces based on locoregional features has improved performance in healthy controls [2]. However, this approach estimates features on vertices arranged without anatomical correspondences, which may ultimately produce inaccurate segmentation in case of topological discrepancies due to pathology.

In patients with TLE, automatic algorithms have generally yielded less satisfactory results than in healthy subjects, with Dice indices ranging from 0.63 to 0.77 and 0.75 to 0.89, respectively [3, 7]. The reduced performance in TLE may stem from factors other than atrophy, as previous studies achieved results similar to controls in patients with Alzheimer’s disease, a neurodegenerative disorder associated with marked hippocampal atrophy [7]. Indeed, 40% of TLE patients show atypical morphologies of the hippocampus and its surroundings, which are referred to as malrotation [8]. In the presence of such abnormal anatomical variant, automatic approaches may fall into local minima when using a template or prior-knowledge based on healthy subjects. Multi-template approaches offer an appropriate framework to account for structural variability by selecting from a database a subset of atlases that best describe anatomical characteristics of the target structure. To date, these techniques have been volume-based and applied to segment healthy hippocampi [4-5]. However, they rely on nonlinear registration of image-intensity that may also fail in case of atypical morphology.

We propose a novel hippocampal segmentation method based on a multi-template approach that statistically models parametric surfaces and locoregional texture features. We applied this method to a large database of healthy controls and patients with TLE.

2 Methods

Our approach TLE consists of a template library construction stage and a segmentation stage, as illustrated in Fig. 1. Each stage is detailed in the following sections.

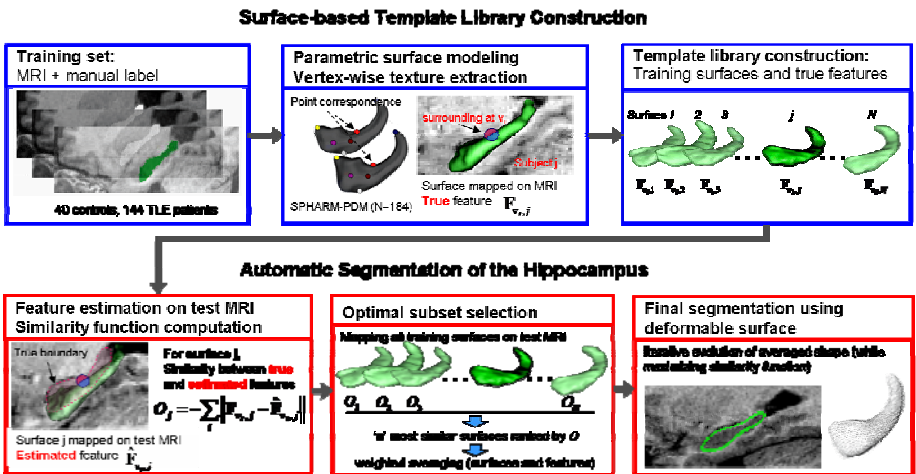


Fig. 1. Flowchart of the proposed segmentation technique

2.1 Template Library Construction

2.1.1 Surface Extraction from Manual Labels

Labels of the hippocampus obtained from manual segmentation were converted into surface meshes using the spherical harmonics parameterization and point distribution models (SPHARM-PDM), an area-preserving, distortion-minimizing spherical mapping that ensures shape-inherent point-wise correspondence through an icosahedron subdivision of the spherical harmonic parameterization [9].

2.1.2 Regional Texture Models

Each surface was mapped to its corresponding image to estimate regional textures. For each hippocampal vertex, a spherical neighborhood with various radii (3mm, 5mm, 7mm) was defined. The “inner region” (IR) and “outer region” (OR) of these local neighborhoods were determined with respect to the surface boundary. The following texture features were then computed at each vertex \mathbf{v}_i .

- i. *Normalized intensity* (NI) to capture regional tissue homogeneity. Let $\mu_{IR, i} / \mu_{OR, i}$ be the mean of intensities within IR or OR at \mathbf{v}_i and $SD_{IR, i} / SD_{OR, i}$ be the standard deviation. We defined $NI_{IR, i} = \mu_{IR, i} / SD_{IR, i}$ and $NI_{OR, i} = \mu_{OR, i} / SD_{OR, i}$.
- ii. *Relative intensity* (RI) to assess the contrast between IR and OR voxels. RI was defined as $RI_i = 2 \times (\mu_{OR, i} - \mu_{IR, i}) / (\mu_{OR, i} + \mu_{IR, i})$.
- iii. *Gabor energy* (GE) to capture image texture through a multi-channel filtering strategy [10]. Mimicking human visual perception, this feature portrays the complexity, directionality and repetition of the intensity distribution. Let x, y, z be the spatial coordinates and \mathbf{R}_θ a 3x3 rotation matrix whose 3D Euler angle θ defines the orientation of the normal to the parallel stripes of a Gabor function. The Gabor filter [10] was defined by:

$$g_{\theta, \lambda, \sigma, \gamma, \varphi}(x, y, z) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2 + \gamma^2 z'^2}{2\sigma^2}\right) \exp\left(-2\pi \frac{x'}{\lambda} + \varphi\right), \quad (x', y', z') = \mathbf{R}_\theta(x, y, z) \quad (1)$$

γ is the aspect constant, φ is the phase offset and the ratio σ/λ describes the bandwidth, namely the filter size b according to:

$$b = \log_2 \frac{\pi\sigma / \lambda + \sqrt{0.5 \ln 2}}{\pi\sigma / \lambda - \sqrt{0.5 \ln 2}}, \quad \sigma / \lambda = \frac{1}{\pi} \sqrt{\frac{\ln 2}{2} \frac{2^b + 1}{2^b - 1}} \quad (2)$$

At voxel $I(x, y, z)$, the Gabor energy was given by $GE_{\theta, \lambda, \sigma, \gamma, \varphi}(x, y, z) = \|g_{\theta, \lambda, \sigma, \gamma, \varphi}(x, y, z) * I(x, y, z)\|$, where $*$ is the convolution operator. In this work, we calculated the Gabor energy of the immediate surroundings only along the surface normal θ . We fixed φ as 0 (no offset) and γ as 1 (same amount of information horizontally and vertically). Multiscale texture analysis was performed by varying the bandwidth $b = \{0.5, 1, 2\}$.

- iv. *Intensity gradient* (IG): to capture edge information. Gradients along x, y, z direction were computed and interpolated on the vertices.

2.1.3 Regional Shape Models

The following features constrained the deformable model evolution within the range of the anatomical variability in the template.

- i. *Distance between adjacent vertices*: to prevent irregular vertex topology.
- ii. *Gaussian curvature*: to constrain local convexity/concavity.
- iii. *Local orientation*: manual tracing in regions where anatomical boundaries are not visible often rely on arbitrary oblique lines as geometric landmarks (e.g., the inferomedial border separating CA1 from the subiculum) [11]. To model this feature, we projected the surface normals to the xy, yz and zx planes and computed their angles with respect to their orthogonal axis x, y and z-axis.

For each subject, texture and shape features were normalized across vertices using a z-transform and combined at each vertex \mathbf{v}_i into a vector $\mathbf{F}_v = [\mathbf{F}_{texture\ v} \ \mathbf{F}_{shape\ v}]$.

2.2 Automatic Segmentation of the Hippocampus

2.2.1 Automatic Selection of an Optimal Shape and Feature Template

Let $\mathbf{S}_j = [\mathbf{v}_1, \mathbf{v}_2 \dots \mathbf{v}_i \dots, \mathbf{v}_L]$ be a SPHARM-PDM surface of the template library, which is initially mapped on its own MR image. Let $\mathbf{F}_{v_i, j}$ be the *true features*, i.e. the set of features computed at \mathbf{v}_i . Given a test image, we mapped the templates $\mathbf{S}_1, \mathbf{S}_2 \dots \mathbf{S}_j \dots \mathbf{S}_N$, to the test image and computed a set of *estimated features* $\hat{\mathbf{F}}_{v_i, j}$. For each mapping, we computed a similarity measure as:

$$O_j = -\sum_i \frac{\|\mathbf{F}_{v_i, j} - \hat{\mathbf{F}}_{v_i, j}\|}{\sqrt{\frac{1}{N} \sum_{k=1}^N (\mathbf{F}_{v_i, k} - \bar{\mathbf{F}}_{v_i})^2}}, \quad \bar{\mathbf{F}}_{v_i} = \frac{1}{N} \sum_{k=1}^N \mathbf{F}_{v_i, k} \quad (3)$$

Equation (3) represents a normalized similarity between the j -th *true features* and *estimated features*. Thus, it enables the selection of the closest template surface. Two approaches were investigated:

- a) Traditional approach: Given a test image, we selected the n most similar templates (surfaces and their corresponding features) [4-5]. We then computed the average surface from these ‘ n ’ surfaces ($\mathbf{S}_{opt\ 1}, \mathbf{S}_{opt\ 2}, \dots, \mathbf{S}_{opt\ n}$) to generate an initial shape for the later segmentation process. Best results were obtained experimentally with $n=10$ templates.
- b) Advanced adaptive approach: Recently, Coupe et al. [12] proposed a weighted averaging of the volume labels in the selected subset according to their similarity measure. Here, we adapted this technique by optimally weighting surfaces and vertex-wise features separately. Let \mathbf{w}_S and \mathbf{w}_F be $n \times 1$ weight vectors for optimal surfaces and features, respectively. We defined the new average surface as:

$$\bar{\mathbf{S}} = \sum_{j=1}^n w_{S,j} \mathbf{S}_{opt j}; \quad \sum w_{S,j} = 1 \quad (4)$$

Similarly, we defined the weighted mean and SD of features at vertex \mathbf{v}_i by:

$$\bar{\mathbf{F}}_{\mathbf{v}_i} = \sum_{j=1}^n w_{F,j} \mathbf{F}_{\mathbf{v}_i, opt j}; \quad \sum w_{F,j} = 1; \quad \sigma_{F \mathbf{v}_i} = \sqrt{\sum_{j=1}^n w_{F,j} (\mathbf{F}_{\mathbf{v}_i, j} - \bar{\mathbf{F}}_{\mathbf{v}_i})^2} \quad (5)$$

We then redefined the similarity function based on the n-top ranked subset as:

$$O_{subset} = - \sum_i \frac{\|\bar{\mathbf{F}}_{\mathbf{v}_i} - \hat{\mathbf{F}}_{\mathbf{v}_i, \bar{s}}\|}{\sigma_F} \quad (6)$$

$\hat{\mathbf{F}}_{\mathbf{v}_i, \bar{s}}$ is the *estimated* feature-set computed on the surface \bar{s} mapped on the test image. Finally, both weights were determined by maximizing the similarity between the current template-subset and the test image.

$$\mathbf{w} = [w_S \ w_F] = \arg \max_{\mathbf{w}} O_{subset} \quad (7)$$

We initially set all the components of \mathbf{w} as 1/n. We then iteratively perturbed every w_j by $\pm\delta \cdot 1/n$ and updated it if the similarity function (7) increased. The step-size parameter δ was initialized to 1 and decreased at each iteration by 0.1.

2.2.2 Automatic Segmentation: Evolution and Objective Function

The final segmentation was obtained as follows. We first linearly mapped the template computed in the previous section to the test image. Then, we locally deformed the surface at each vertex along the surface normal based on a multi-level b-spline interpolation. Analogous to Eq. (6), the cost function was defined at iteration k as:

$$O_k = - \sum_i \frac{\|\bar{\mathbf{F}}_{\mathbf{v}_i} - \hat{\mathbf{F}}_{\mathbf{v}_i, \bar{s}_k}\|}{\sigma_F} \quad (8)$$

\bar{s}_k is a deformed surface at iteration k and $\hat{\mathbf{F}}_{\mathbf{v}_i, \bar{s}_k}$ is its *estimated* feature vector. We maximized O_k using the gradient descent approach.

3 Experiments and Results

3.1 Experiments

Our training-set included 40 healthy controls (18 males; mean age 33 ± 12 yrs) and 144 TLE patients (61 males; mean age 36 ± 11 yrs). TLE diagnosis and lateralization of the seizure focus into left TLE (LTLE; n=73) and right TLE (RTLE; n=71) were determined by a comprehensive evaluation including video-EEG recordings and MRI evaluation. Our ethics committee approved the study. MR images were acquired on a 1.5T scanner, with 1mm-isotropic voxels. All images underwent automated correction for intensity non-uniformity, intensity standardization, and were linearly registered

into a stereotaxic space [13]. The hippocampus was manually segmented by an expert in all subjects. Using z-score normalization based on the distribution of healthy controls, we identified 91 (63%) patients with hippocampal atrophy (i.e. $z < -2$) ipsilateral to the seizure focus.

3.1.1 Evaluation of Template Selection Approaches

We evaluated the proposed surface-based multi-template methods (SurfMulti) through a leave-one out strategy. For each test data, we selected optimal subsets and built the initial shapes and feature models using both the traditional approach and the adaptive approach. The segmentations resulting from each initialization were compared using Dice index and paired t-test.

3.1.2 Comparison with Volume-Based Single- and Multi-template Approaches

We obtained hippocampal segmentation using FreeSurfer, a volume-based single-template approach [6], and a multi-template approach [4-5] (that we named VolMulti). For VolMulti, we used ANIMAL as the nonlinear registration method [13] and chose the optimal subset of 11 as suggested by [4].

We evaluated the performance of each automated segmentation algorithm against the manual label using the Dice index. We compared Dice indices between automated methods in controls and each patient group (i.e. hippocampi ipsilateral / contralateral to the seizure focus) using Student's t-tests.

In a separate analysis, we assessed the sensitivity of each algorithm to detect atrophy in TLE relative to controls by computing Cohen's $d = (\text{mean volume controls} - \text{mean volume TLE}) / \text{pooled SD}$, indicating the effect size of a 'between group' difference. We computed the significance of the observed effect using t-tests.

Significances of all statistical tests were adjusted for multiple comparisons using Bonferroni-correction.

3.2 Results

The adaptive template selection approach outperformed the traditional strategy (all groups: $p < 0.02$, Table 1). We therefore took as reference the adaptive approach.

In all groups, performance of SurfMulti was superior to the two volume-based approaches (vs. FreeSurfer: $p < 10e-15$; vs. VolMulti: $p < 0.0004$). Moreover, our algorithm performed equally well in TLE patients and controls ($p > 0.1$), whereas volume-based approaches segmented poorly hippocampi in patients (FreeSurfer: $p < 10e-6$, VolMulti: $p = 0.02$).

Group-wise comparisons identified hippocampal atrophy ipsilateral to the seizure focus in TLE patients irrespective of the method, i.e. manual or automated ($p < 0.05$, Table 2). However, the effect size of atrophy detected using SurfMulti was closest to manual labeling (Cohen's d : Manual=1.71, $t=7.6$; SurfMulti=1.60, $t=7.0$; VolMulti=1.38, $t=6.1$; FreeSurfer=0.91, $t=3.9$).

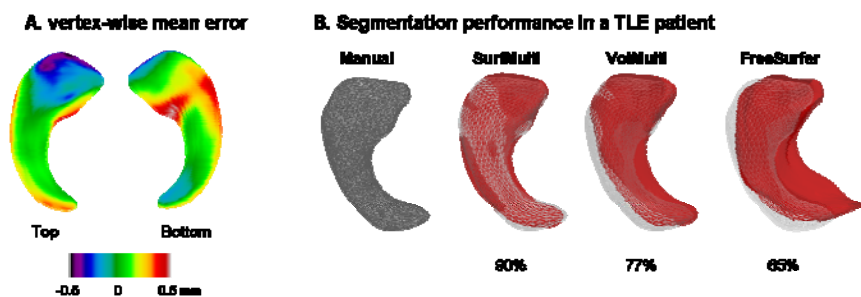
For each test data, we computed vertex-wise displacements between the SPHARM-PDM surfaces of ground truth and SurfMulti segmentations. Fig. 2-A illustrates the mean and SD of the normal displacement across subjects ($n=184$). This surface-based analysis showed that SurfMulti yielded results with sub-voxel error (absolute mean error < 0.51 mm; mean SD=0.9mm).

Table 1. Segmentation accuracy as defined by Dice similarity index (% mean \pm SD)

Group	FreeSurfer	VolMulti	SurfMulti	
			Equal weight	Adaptive weight
<i>Controls</i>	72.1 \pm 2.5	84.5 \pm 4.5	86.4 \pm 3.2	87.5 \pm 2.6
<i>Ipsilateral</i>	65.8 \pm 4.5	81.2 \pm 5.7	84.5 \pm 3.3	86.1 \pm 2.9
<i>Contralateral</i>	70.0 \pm 3.7	82.7 \pm 5.5	86.6 \pm 3.5	87.7 \pm 2.7

Table 2. Hippocampal volumetry. Values represent z-scores with respect to controls (mean \pm SD) and the effect size of atrophy as Cohen's d (within brackets; 0.2 indicates a small effect, 0.5 a medium effect, and >0.8 is a large effect).

	Manual	FreeSurfer	VolMulti	SurfMulti
<i>Ipsilateral</i>	-2.41 \pm 1.85 (1.71)	-1.35 \pm 2.06 (0.91)	-1.49 \pm 1.40 (1.38)	-1.78 \pm 1.45 (1.60)
<i>Contralateral</i>	-0.65 \pm 1.40 (0.56)	-0.05 \pm 1.91 (0.03)	-0.22 \pm 1.29 (0.22)	-0.57 \pm 1.41 (0.47)

**Fig. 2.** Vertex-wise comparison between manual (ground-truth) and automated segmentation. (A) Mean segmentation error of SurfMulti across all subjects (n=184) mapped on the template. (B) Representative example of a TLE patient showing the overlap between manual labeling (wireframe) of the left hippocampus and the three automatic segmentation methods (in red). The Dice index is indicated below each method.

4 Discussion and Conclusion

We propose a novel hippocampal segmentation algorithm that integrates deformable parametric surfaces and multiple templates in a unified framework. Our technique achieved a level of accuracy in TLE patients virtually identical to healthy controls, with a Dice index of 86.1% for the atrophic hippocampus ipsilateral to the seizure focus. Such performance has not yet been paralleled in epilepsy. Vertex-wise surface-based shape mapping showed that SurfMulti with adaptive weight had an excellent overlap with manual labels, with sub-millimetric precision. Furthermore, we achieved the same sensitivity than manual volumetry in detecting atrophy ipsilateral to the focus. Given that the presence of hippocampal atrophy in TLE predicts a favorable

seizure outcome after surgery, the proposed technique assures to be a robust surrogate tool for the time-demanding manual procedure in the presurgical evaluation.

To segment brain structures, multi-template algorithms use either mutual information [4] or entropy [5] as a measure of intensity similarity between the target and template. In these methods, if intensity distributions of the two images are different, segmentation may fail. Our features, on the contrary, capture intrinsic image characteristics by computing higher order semantic features from a given image, i.e. image homogeneity, contrast, gradient and texture. Moreover, features are weighted according to both mean and SD of the optimal subset.

Although the volume-based multi-atlas method performed globally better than the single-template approach, its agreement with manual segmentation was significantly lower in patients than controls. On the other hand, our method outperformed both state-of-the-art algorithms in both controls and patients and sensitivity in detecting atrophy. This likely results from the integration of: *i*) surface-based shape-inherent point-wise correspondences guaranteed by SPHARM-PDM; *ii*) vertex-wise sampling scheme with respect to the surface boundary, allowing for a better characterization of locoregional texture and shape of structures neighboring the hippocampus; and *iii*) multi-template library derived from a large cohort of healthy controls and patients, that accounts for inter-subject variability, in particular shape variants.

References

1. Schramm, J., Clusmann, H.: The surgery of epilepsy. *Neurosurgery* 62 S2, 463–481; discussion 481 (2008)
2. Pitiot, A., Delingette, H., Thompson, P.M., Ayache, N.: Expert knowledge-guided segmentation system for brain mri. *NeuroImage* 23, S85–S96 (2004)
3. Avants, B.B., Yushkevich, P., Pluta, J., Minkoff, D., Korczykowski, M., Detre, J., Gee, J.C.: The optimal template effect in hippocampus studies of diseased populations. *Neuroimage* 49(3), 2457–2466 (2010)
4. Collins, D.L., Pruessner, J.C.: Towards accurate, automatic segmentation of the hippocampus and amygdala from mri by augmenting animal with a template library and label fusion. *Neuroimage* 52(4), 1355–1366 (2010)
5. Aljabar, P., Heckemann, R.A., Hammers, A., Hajnal, J.V., Rueckert, D.: Multi-atlas based segmentation of brain images. *NeuroImage* 46, 726–738 (2009)
6. Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der, K.A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., Dale, A.M.: Whole brain segmentation. Automated labeling of neuroanatomical structures in the human brain. *Neuron*. 33(3), 341–355 (2002)
7. Chupin, M., Hammers, A., Liu, R.S., Colliot, O., Burdett, J., Bardinnet, E., Duncan, J.S., Garnero, L., Lemieux, L.: Automatic segmentation of the hippocampus and the amygdala driven by hybrid constraints. *NeuroImage* (2009)
8. Bernasconi, N., Kinay, D., Andermann, F., Antel, S., Bernasconi, A.: Analysis of shape and positioning of the hippocampal formation: An mri study in patients with partial epilepsy and healthy controls. *Brain* 128(Pt 10), 2442–2452 (2005)
9. Styner, M., Oguz, I., Xu, S., Brechbühler, C., Pantazis, D., Gerig, G.: Statistical shape analysis of brain structures using spharm-pdm. In: *MICCAI 2006 Opensource Workshop* (2006)

10. Grigorescu, S.E., Petkov, N., Kruizinga, P.: Comparison of texture features based on gabor filters. *IEEE Trans. Image Process.* 11(10), 1160–1167 (2002)
11. Konrad, C., Ukas, T., Nebel, C., Arolt, V., Toga, A.W., Narr, K.L.: Defining the human hippocampus in cerebral magnetic resonance images—an overview of current segmentation protocols. *Neuroimage* 47(4), 1185–1195 (2009)
12. Coupé, P., Manjón, J.V., Fonov, V., Pruessner, J., Robles, M., Collins, D.L.: Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation. *Neuroimage* 54(2), 940–954 (2011)
13. Collins, D.L., Neelin, P., Peters, T.M., Evans, A.C.: Automatic 3d intersubject registration of mr volumetric data in standardized talairach space. *J. Comput. Assist. Tomogr.* 18(2), 192–205 (1994)

Model Based 3D Segmentation and OCT Image Undistortion of Percutaneous Implants

Oliver Müller¹, Sabine Donner^{2,3}, Tobias Klinder⁴, Ralf Dragon¹,
Ivonne Bartsch³, Frank Witte³, Alexander Krüger^{2,3},
Alexander Heisterkamp^{2,3}, and Bodo Rosenhahn^{1,*}

¹ Institut für Informationsverarbeitung, Leibniz Universität Hannover,
Appelstraße 9a, 30167 Hannover, Germany

{omuller, rosenhahn}@tnt.uni-hannover.de

² Laser Zentrum Hannover e.V., Hollerithallee 8, 30419 Hannover, Germany

³ CrossBIT, Center for Biocompatibility and Implant-Immunology,
Hannover Medical School, Feodor-Lynen-Straße 31, 30625 Hannover, Germany

⁴ Philips Research North America, Briarcliff Manor, NY 10510, USA

Abstract. Optical Coherence Tomography (OCT) is a noninvasive imaging technique which is used here for in vivo biocompatibility studies of percutaneous implants. A prerequisite for a morphometric analysis of the OCT images is the correction of optical distortions caused by the index of refraction in the tissue. We propose a fully automatic approach for 3D segmentation of percutaneous implants using Markov random fields. Refraction correction is done by using the subcutaneous implant base as a prior for model based estimation of the refractive index using a generalized Hough transform. Experiments show the competitiveness of our algorithm towards manual segmentations done by experts.

1 Introduction

Optical Coherence Tomography is a non invasive imaging modality used for taking optical biopsies of layered tissue structures such as the epidermis [11] and the retina [4,14]. Apart from clinical use, OCT also has applications in animal studies with the advantage of repetitive biopsies at one animal at different time points instead of lethal biopsies at different animals for each time point. The particular objective of this study is the morphometric analysis of the skin in the vicinity of a percutaneous implant situated in the lateral abdominal region of a hairless mouse to draw conclusions on its biocompatibility (see Fig. 1(a)).

As the optical properties of the tissue introduce distortions into the OCT images [13], segmentation based image undistortion is an important step towards fully automatic image analysis tasks. In recent works, several methods like graph based global optimization, active contours and random fields are proposed for layer segmentation. In practice, graph based approaches, as used in [5] for fully

* This work has been partially funded by the DFG within the excellence cluster RE-BIRTH.

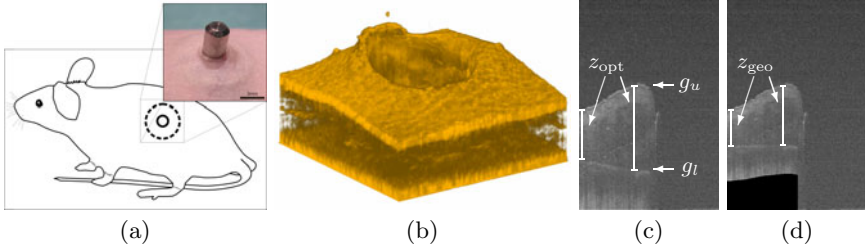


Fig. 1. (a) Location and photo of a percutaneous implant, (b) OCT dense 3D scan volume rendering (percutaneous pin is not visible), (c) single OCT B-scan (cropped at half) showing the distorted baseline, and (d) corresponding undistorted result

automatic 3D retinal multilayer segmentation, lead to huge graph sets, limiting the number of voxels. Active contour models (e.g. snakes) [9,14] provide robust results, however require manual initialization. In [8], a fully automatic 2D feature segmentation is presented using conditional random fields and efficient optimization algorithms for inference. Segmentation of a single 2D OCT scan (B-scan) can be susceptible to local shading effects and image perturbations and extending the scanning scheme to the third dimension can significantly improve the segmentation quality [3,5,6]. Thus our algorithm is based on 3D segmentation.

This paper proposes an approach for fully automatic segmentation of 3D Fourier-domain OCT and refractive undistortion. The determination of the refractive index is facilitated by the geometry of the implant which consists of a percutaneous pin (3 mm diameter and 5 mm length) anchored beneath the dermis by a flat disc shaped base which is visible in OCT (see Fig. 1(a)-(d)).

Two main technical **contributions** are proposed. First, estimation of the skin surface in the 3D space from several OCT B-scans is done using a Markov random field (MRF) approach with an efficient combination of global and local optimization algorithms. A spoke pattern scanning scheme is used for 3D data acquisition and is further compared with a dense 3D scanning scheme (see Fig. 2(a)). Our second contribution addresses the segmentation of the implant base. The distorted implant base is segmented using a refractive distortion model and the previously segmented skin surface for parameter estimation in order to match the distorted implant base best to the a priori known shape of the undistorted base contour. The parameters of the implant base are estimated with a fast generalized 3D Hough transform approach, optimizing the refractive index, as well as the 3D position and orientation of the implant base. The segmented model is finally used for refractive image undistortion (see Fig. 1(c)-(d)).

In Section 2, the 3D segmentation of the pin position, the skin surface and the implant base is described (Fig. 2(b)), which is used for refractive image undistortion (Fig. 2(c)). In Section 3, the used undistortion model is verified and a comparison of the spoke pattern and dense 3D scanning scheme is shown, followed by a quantitative analysis of several mouse datasets segmented and undistorted using our method. Finally, in Section 4, a short conclusion is given.

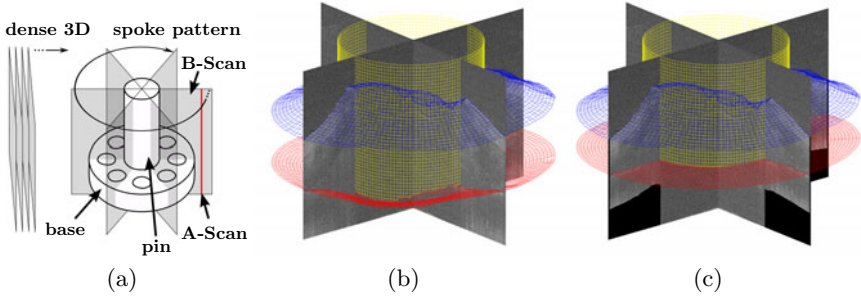


Fig. 2. (a) Schematic of the percutaneous implant (pin and base) and OCT scanning schemes (dense 3D and spoke pattern), (b) two orthogonal OCT B-scans with segmentation (mesh overlay) of pin position (yellow), skin surface (blue) and deformed base (red), (c) undistorted B-scans with mesh overlay

2 Methods

The OCT data is acquired in a sequence of B-scans $(I^k)_{k=1,\dots,K}$ (see Fig. 2(a)) with image width W and height H . To reduce noise and small scanning artifacts, while preserving edges, we apply a median filter to each B-scan as a first preprocessing step. In a second step, we apply a pixel intensity normalization to each image, leading to a zero-mean intensity distribution with unit variance, in order to retrieve uniform edge responses from the later applied edge filter. We use a Sobel filter in combination with a presmoothing Gaussian kernel with $\sigma_{\text{gauss}} = 1.5$ to get first order derivative images I_x^k and I_y^k in x- and y-direction of I^k .

For 3D segmentation, the two-dimensional B-scans are embedded in a global 3D coordinate system. A mapping of a 3D point position $\mathbf{P} = (X, Y, Z)$ into image coordinates $\mathbf{p} = (x, y)$, i.e. $(X, Y, Z) \mapsto (k, x, y)$ as shown in Fig. 3(a) is done, resulting in a sparse volume representation $V(X, Y, Z) = I^k(x, y)$ for the image intensities. The volumes of the image derivatives V_x , and V_y are analogously defined, using I_x^k and I_y^k instead of I^k .

The proposed implant segmentation method is divided into three consecutive steps (see Fig. 2(b)-2(c)): The pin segmentation (yellow cylinder), the skin surface segmentation (blue mesh), and the base segmentation (red mesh).

2.1 Pin Segmentation

In the image area where the implant pin is located, there is no tissue, and hence no contour information. Therefore, this area can be ignored for following segmentation steps. The pin is of cylindrical form and the diameter d is known a priori. The diameter is allowed to have a variance v^k of ± 10 pel. For each image I^k , the x-position of the left and right pin boundary x_1^k and $x_2^k = x_1^k + d + v^k$ is computed using a generalized Hough transform [1] approach over $I_{\text{acc}}^k(x) = \sum_y |I_x^k(x, y)|$

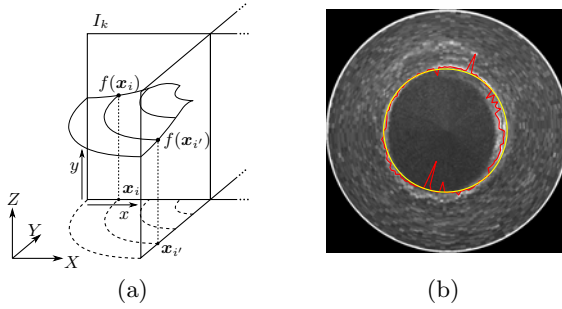


Fig. 3. (a) Modeling of skin surface f over grid points $x_i, x_{i'}$ with respect to global coordinate system (X, Y, Z) using spoke pattern scanning scheme, (b) top view of accumulated image intensities I_{acc} and segmented pin boundary for each B-scan individually (red), and with fitted cylinder (yellow)

with $\max_{x_1^k, v^k} [I_{acc}^k(x_1^k) + I_{acc}^k(x_1^k + d + v^k)]$. The pin estimation in a single B-scan is susceptible to noise and vanishing of boundary contours (see Fig. 3(b), red line). Therefore, the boundaries are smoothed by assuming a cylindrical form and radius r of the pin: At first, the center position $C = (C_X, C_Y)$ of the pin in the XY -plane is calculated as the arithmetic mean of all boundary positions B_1^k and B_2^k corresponding to x_1^k and x_2^k , i.e. $C = \frac{1}{2K} \sum_k (B_1^k + B_2^k)$. Then, the radius r is computed as the median of all radii $r = \text{median}(|CB_1^1|, \dots, |CB_1^K|)$. The yellow line in Fig. 3(b) shows a typical result of the pin segmentation step.

2.2 Skin Modeling and Segmentation

A correct estimation of the skin surface is crucial for a correct modeling of the implant base. The skin surface, denoted as $f(X, Y) = Z$, is assumed to be smooth and to behave like a membrane, thus having no discontinuities, except the pinhole. Several approaches can be used to model the surface: Markov Random Fields (MRF), Conditional Random Fields or Discriminative Random Fields [10]. Due to the smoothness property, we decided to use an adapted MRF for segmentation of the skin surface. Following the notations in [10], the posterior probability of the skin surface f given the volumetric pixel intensities V can be written as $P(f | V) \propto P(f, V) = p(V | f)P(f)$ using the Maximum A-Posteriori framework (MAP), where $P(f)$ is the smoothness prior and $p(V | f)$ denotes the likelihood function of f for V fixed.

To represent probabilistic relationships of the MRF, a common graphical notation is used. The neighborhood relationships of a MRF can be described given a graph $G = (V, E)$ with a set of nodes V representing the instances of the random field and a set of edges E representing the conditional dependencies between the instances. The prior probability is then modeled as follows:

$$P(f) = \exp \left[- \sum_{i \in V} U(f_i) \right] \quad \text{with} \quad U(f_i) = \sum_{\{i, i'\} \in E} (f_i - f_{i'})^2 / 2\sigma_s^2, \quad (1)$$

where σ_s is a constant weighting factor. Figure 3(a) shows the relation of the skin surface $f_i = f(\mathbf{x}_i)$ in the observation point \mathbf{x}_i to its neighboring point \mathbf{x}'_i . A 4-neighborhood system is used.

The likelihood of the true skin surface at position $\mathbf{x}_i = (x_i, y_i)$ going through the volumetric point (X_i, Y_i, Z_i) , with $Z_i = f(\mathbf{x}_i)$ is given as

$$p(V|f_i) \propto V_y(X_i, Y_i, f(\mathbf{x}_i)) + c, \quad (2)$$

with a shifting constant c , forcing strict positive probabilities. Determining the optimal solution of the given MRF problem is transferred to finding the global maximum of $P(f | V)$. The given MRF model consists of $K \cdot W$ edges and $K \cdot W \cdot H$ observation points, thus searching for the global maximum of the joint probability turns out to be a complex task. In order to solve this task in a reasonable amount of time and memory usage, we decided to use an iterative local optimization algorithm. The Iterated Conditional Modes (ICM) approach with the *coding method* of Besag [2] is used because of its ability for fast convergence. The ICM algorithm searches for a local maximum of the joint probability $P(f | V)$ by iteratively maximizing each local probability $P(f_i | V)$ independently:

$$f_i^{n+1} = \arg \min_z [-\log p(V | f_i^n = z) + \sum_{\{i, i'\} \in E} (z - f_{i'}^n)^2 / 2\sigma_s^2], \quad (3)$$

where f^0 is an initial guess of the surface. ICM is a local minimization method and the estimation result highly depends on the initial surface guess f^0 . Therefore, the initial guess f^0 is retrieved by independently estimating an optimal path for each B-scan using a Markov model, i.e., the same model as for the MRF, but with a 2- instead of a 4-neighborhood system. Finally, the Viterbi algorithm is used for global optimization [12]. Additionally, an annealing procedure inspired by annealing labeling ICM of [10] is used, i.e. allowing the membrane for $n = 0$ to be more relaxed by setting σ_s^0 to a higher value σ_{start} and decreasing it for each L 's iteration by $\sigma_s^{n+L} = \max\{\sigma_{\text{end}}, \sigma_s^n \cdot \sigma_{\text{decr}}\}$.

2.3 Baseline Modeling and Detection with a Hough Transformation

To achieve an appropriate segmentation of the implant base in presence of scanning artifacts, noise, and local vanishing edge structures, a robust and model based segmentation approach was developed using a generalized Hough transform [1]. The applied refractive image undistortion model uses the fact that the axial position of reflections captured with the OCT system matches to the optical path length z_{opt} of light passing through the observed tissue, rather than the geometric path length z_{geo} . Inspired by [11], the relation between z_{opt} and z_{geo} can be approximately formulated as $z_{\text{opt}} = nz_{\text{geo}}$, as shown in Fig. 1(c)–(d), assuming a homogeneous layer with refractive index n . The model used for conversion of optical path length to geometric path length of each axial scan (A-scan, see Fig. 2(a)) is $z = g_u + (g_l - g_u)/n$, where g_u is the known upper position (the skin surface, estimated in Section 2.2), g_l the lower position (base layer), and z denotes the geometric position of the base layer (see Fig. 1(c)).

Given g_u and the constraint of maximal edge intensity support of I_y over g_l , the maximization term for the generalized Hough transform is stated as follows:

$$\max_{Z, n, \theta_X, \theta_Y} \sum_{x, k} I_{y, k}(x, g_u(x, k) - [g_u(x, k) - Z_{\theta_X, \theta_Y}(x, k)] \cdot n) . \quad (4)$$

The implant base is modeled as a plane with geometric position Z . Since the pin is not located perfectly horizontal, a rotation of the plane in X - and Y -direction is applied to Z , denoted by Z_{θ_X, θ_Y} . As the parameter space is of dimension 4, small discretization step sizes lead to high computation time, i.e. doubling the precision increases the computation time by a factor of 2^4 . Since the minimization problem has only one global optimum, which can be distinguished very well from small local extrema, a resolution pyramid approach [7] is applied.

3 Experiments

In this section, a ground truth experiment for verification of the proposed undistortion model, as well as a comparison of the used spoke pattern and dense 3D scanning scheme is performed. We further show results of a quantitative study on several mouse datasets¹. The B-scans have a dimension of $800 \text{ px} \times 600 \text{ px}$ with a lateral distance of the A-Scans of $7.5 \mu\text{m}/\text{px}$ and an axial scale of $4.7 \mu\text{m}/\text{px}$. Following parameters work best for our datasets: An 11×11 median filter, $\sigma_{\text{start}} = 70$, $\sigma_{\text{end}} = 10$, $\sigma_{\text{dec}} = 0.9$, and $L = 5$.

Model Verification: A plane plastic slide is prepared with two glue drops of slightly different size (Vitalit[®] 4731) with known refractive index of $n = 1.474$. An example image and its segmentation results are shown in Fig. 4. The estimated refractive index of the two glue drops are $n_{\text{est}} = 1.494$ and $n_{\text{est}} = 1.507$ respectively. It is assumed that the ground carrier plate is not perfectly planar as expected by the estimation model, causing the deviation from groundtruth.

Scanning Scheme: The spoke pattern scanning scheme is compared with the dense 3D scanning scheme using real mouse recordings. To this end, we use the skin surface reconstructions of scans captured using the spoke pattern with 72 B-scans and the dense 3D scanning scheme, respectively. Figure 5(a) shows reconstructions for a typical skin surface (acquired post mortem). The segmented surface using the spoke pattern is projected onto the dense 3D grid. Small surface differences show the ability of capturing important surface features using the spoke pattern (see Fig. 5(b)). The root mean square deviation (RMSD) is calculated as 1.546 pel. Furthermore, several other surfaces were reconstructed (see Fig. 5(c)). The results show, that using the spoke pattern still leads to good results and preserves many skin details. Moreover, scanning is faster and approximately 91 % of computation time and disk space is saved due to the decreased

¹ All animal experimental procedures have been approved by the local governmental animal care committee (Approval No. 33-42502-04-08/1498).

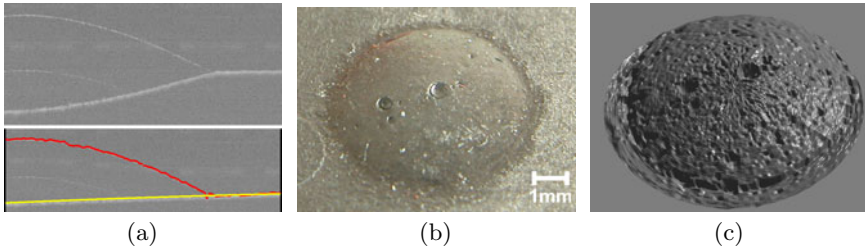


Fig. 4. Glue drop (Vitrilit[®] 4731) with $n = 1.474$, (a) comparison of original (top) with segmented and undistorted (bottom) OCT B-scan (images cropped), (b) closeup photo of a glue drop, (c) corresponding rendered surface reconstruction

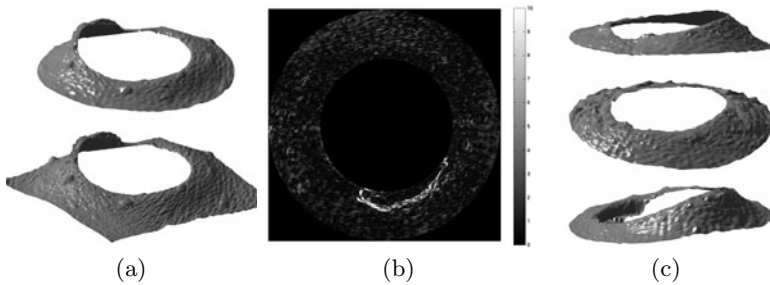


Fig. 5. (a) Comparison of spoke pattern scanning scheme (top) and dense 3D scanning scheme (bottom), (b) comparison heightmap of absolute surface differences in pel, (c) example skin segmentation results of mouse datasets using spoke pattern

number of B-scans. With 937.5 KiB per image, 666.5 MiB are saved. Skin surfaces of spoke pattern scans are reconstructed using an unoptimized MATLAB implementation in 5.78 min, compared to 68.14 min using dense 3D scans.

Quantitative Study: We further carried out a quantitative analysis on a set of 60 OCT mouse scans of 23 mice at various points of time (including post mortem scans) using the spoke pattern scanning scheme. Manual segmentations of the skin and base were done from experts for 8 OCT images (each 9th slice) per scan, having 3 individual manual segmentations per slice. The experts were instructed to trace only visible parts of the contours. As a metric, we use the RMSD of a surface S_1 towards the mean of a set of surfaces S_2, \dots, S_m . For each B-scan, the RMSD of the automatic skin segmentation towards the mean of all manual skin segmentations is calculated. The average and standard deviation (in pel) over all B-scans is 3.98 ± 3.29 . For comparison, the RMSD of each manual skin segmentation towards the mean of all other manual skin segmentations is found as 3.62 ± 1.03 . For automatic vs. manual base segmentations, we get: 12.90 ± 18.27 and 3.55 ± 4.51 . After outlier removal (B-scans with $\text{RMSD} \geq 10$ pel), the RMSD of our fully automated approach is 3.63 ± 1.33 (2.3% outlier)

and 5.27 ± 2.40 (39.0% outlier), which is close to the RMSD of the manual segmentation performed by experts with 3.62 ± 1.00 (0.1% outlier) and 3.21 ± 1.37 (1.4% outlier). Outliers are mostly due to motion artifacts in scans captured in vivo. Future work will concentrate on reducing the outliers.

4 Conclusion

In this paper we propose a fully automatic approach for 3D segmentation of percutaneous implants using Markov random fields with application to refractive image undistortion. The refraction correction is done by using the subcutaneous implant baseline as a prior for model based estimation of the refractive index using a generalized Hough transform. Several experiments on the undistortion model including a quantitative evaluation show the competitiveness of our algorithm compared to manual segmentation by experts.

References

1. Ballard, D.H.: Generalizing the hough transform to detect arbitrary shapes. *Pattern Recognition* 13(2), 111–122 (1981)
2. Besag, J.: Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society, B* 36, 192–236 (1974)
3. Eichel, J.A., Bizheva, K.K., Clausi, D.A., Fieguth, P.W.: Automated 3D Reconstruction and Segmentation from Optical Coherence Tomography. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) *ECCV 2010, Part III*. LNCS, vol. 6313, pp. 44–57. Springer, Heidelberg (2010)
4. Fernández, D.C., Salinas, H.M., Puliafito, C.A.: Automated detection of retinal layer structures on optical coherence tomography images. *Optics Express* 13(25), 10200–10216 (2005)
5. Garvin, M.K., Abrámoff, M.D., Kardon, R., Russell, S.R., Wu, X., Sonka, M.: Intraretinal Layer Segmentation of Macular Optical Coherence Tomography Images Using Optimal 3-D Graph Search. *IEEE Transactions on Medical Imaging* 27(10), 1495–1505 (2008)
6. Hori, Y., Yasuno, Y.: Automatic Characterization and Segmentation of Human Skin using Three-Dimensional Optical Coherence Tomography. *Optics Express* 14(5), 1862–1877 (2006)
7. Illingworth, J., Kittler, J.: A Survey of the Hough Transform. *Computer Vision, Graphics, and Image Processing* 44, 87–116 (1988)
8. Karimaghloo, Z., Shah, M., Francis, S.J., Arnold, D.L., Collins, D.L., Arbel, T.: Detection of Gad-Enhancing Lesions in Multiple Sclerosis Using Conditional Random Fields. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) *MIC-CAI 2010*. LNCS, vol. 6363, pp. 41–48. Springer, Heidelberg (2010)
9. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active Contour Models. *International Journal of Computer Vision* 1(4), 321–331 (1988)
10. Li, S.Z.: *Markov Random Field Modeling in Image Analysis*, 3 edn. Springer Publishing Company, Incorporated (2009)
11. Tearney, G.J., Brezinski, M.E., Southern, J.F., Bouma, B.E., Hee, M.R., Fujimoto, J.G.: Determination of the refractive index of highly scattering human tissue by optical coherence tomography. *Optics Letters* 20(21), 2258–2260 (1995)

12. Viterbi, A.J.: Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm. *IEEE Trans. on Inform. Theory* 13(2), 260–269 (1967)
13. Westphal, V., Rollins, A.M., Radhakrishnan, S.: Correction of geometric and refractive image distortions in optical coherence tomography applying Fermats principle. *Optics Express* 10(9), 397–404 (2002)
14. Yazdanpanah, A., Hamarneh, G., Smith, B., Sarunic, M.: Intra-retinal Layer Segmentation in Optical Coherence Tomography Using an Active Contour Approach. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) *MICCAI 2009, Part II*. LNCS, vol. 5762, pp. 649–656. Springer, Heidelberg (2009)

Optimizing GHT-Based Heart Localization in an Automatic Segmentation Chain

Axel Saalbach¹, Irina Wächter-Stehle¹, Reinhard Kneser²,
Sabine Mollus², Jochen Peters¹, and Jürgen Weese¹

¹ Philips Research, Röntgenstrasse 24 - 26, 22335 Hamburg, Germany

² Philips Research, Weissshausstrasse 2, 52066 Aachen, Germany
axel.saalbach@philips.com

Abstract. With automated image analysis tools entering rapidly the clinical practice, the demands regarding reliability, accuracy, and speed are strongly increasing. Systematic testing approaches to determine optimal parameter settings and to select algorithm design variants become essential in this context. We present an approach to optimize organ localization in a complex segmentation chain consisting of organ localization, parametric organ model adaptation, and deformable adaptation. In particular, we consider the Generalized Hough Transformation (GHT) and 3D heart segmentation in Computed Tomography Angiography (CTA) images. We rate the performance of our GHT variant by the initialization error and by computation time. Systematic parameter testing on a compute cluster allows to identify a parametrization with a good trade-off between reliability and speed. This is achieved with coarse image sampling, a coarse Hough space resolution and a filtering step that we introduced to remove unspecific edges. Finally we show that optimization of the GHT parametrization results in a segmentation chain with reduced failure rates.

Keywords: GHT, heart localization, heart segmentation, CTA.

1 Introduction

It is well known that the performance of image processing and analysis algorithms in terms of reliability, accuracy, and speed can strongly depend on parameter settings and design variants. Systematic approaches to determine optimal parameter settings and algorithm design variants are one approach to address this challenge. In general, the algorithm performance can be optimized by systematically testing parameter value combinations on image databases and selecting those values that minimize a metric of the processing result as used during validation. Hautvast et al. [1], for instance, performed full factorial experiments and analyzed the results using the technique of analysis of variances (ANOVA) to optimize parameter settings of their cardiac MR contour segmentation.

Complex image processing algorithms such as fully automatic heart segmentation [2,3] depend on many design decisions and parameter settings, making

straightforward optimization difficult. On the one hand, due to processing time demands, an extensive optimization of an algorithm is often not feasible, even when large computing resources are available. Tools from experimental design [4] can help to efficiently identify relevant parameter combinations and find optimal parameter values with limited effort. On the other hand, the influence of a single parameter or design variant may have a negligible influence on a metric characterizing the overall processing result. Therefore, approaches for the optimization of individual parts of a complex processing chain are important. An example is Simulated Search [5] allowing to select optimal boundary detection functions per triangle in model-based segmentation.

We present an approach to optimize the initial organ localization step in a complex segmentation chain consisting of organ localization, parametric organ model adaptation, and deformable adaptation. In particular we consider the Generalized Hough Transformation (GHT) [6] for 3D localization and model-based heart segmentation in CTA images. We show that reliability and computational speed of our GHT variant can be optimized by systematic parameter testing using a metric for the initialization error resulting in an improved overall segmentation chain.

In the following section, the segmentation chain is outlined, our GHT variant is described and parameter optimization is introduced. Section 3 describes the experiments and section 4 presents our conclusions.

2 Methods

2.1 Segmentation Chain

Automatic heart segmentation can be done in four steps. First, the heart is localized with a GHT and the heart model is placed in the image. Second, pose and scaling of the heart model are adjusted by parametric adaptation. Third, the relative sizes of the heart chambers are adjusted using a parametric adaptation with a piece-wise affine transformation. Finally, accurate adaptation is performed using shape constrained deformable models.

Fig. 1 shows how adaptation of the heart model to the image improves in the individual steps. All steps are important, because each step relies on a sufficiently accurate initialization provided by the previous step [2,5]. Initial heart localization is very crucial as the entire segmentation chain fails, if heart localization fails or is too inaccurate.

2.2 GHT Algorithm

The GHT [6] is a robust and powerful method for detecting a pre-defined shape undergoing geometric transformations in an image. During learning, a description of the target shape is encoded into a reference table also known as R-table. The entries of the R-table are vectors pointing from shape boundary points to a reference point. These vectors are binned by the orientation of the boundary point. For detecting the position of the shape in a new image, edges are determined. For each detected edge element, the gradient orientation is measured

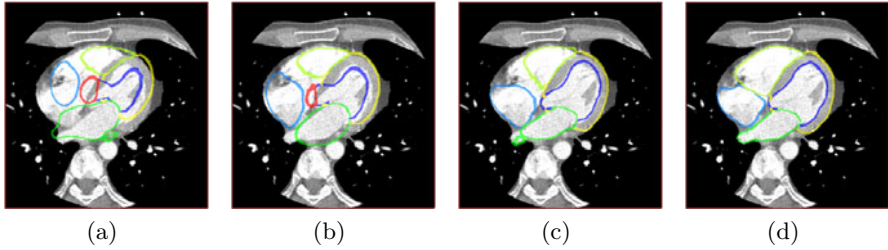


Fig. 1. Heart mesh after GHT-based localization (a), parametric adaptation with a similarity (b) and piecewise-affine (c) transformation and after deformable adaptation (d) overlaid onto a CTA image.

yielding an index for an R-table entry, and the positions pointed by all vectors under this entry are incremented in an accumulator array. This process is called voting and the accumulator array is also referred to as Hough space. The final shape location is given by the highest peak in the accumulator array. Next to translations, further degrees of freedom such as rotation and scaling can be handled by increasing the dimension of the Hough space.

Our GHT implementation can be considered as a 3D variant of the Shape Variant Hough Transform proposed by Brejl and Sonka [7], while we focus in this application on the determination of translations and scaling levels. In order to take variations in heart shape and orientation into account, images and corresponding reference meshes of N_{train} individuals were used for training. The center of gravity of the individual meshes is used to establish a common reference point. Fig. 2 shows the heart model and a flow chart of our GHT algorithm which offers multiple options and parameter settings as described in the following.

For our GHT algorithm, training starts by estimating the image gradient orientation at the triangle centers of the N_{train} heart meshes, which can be done with a $3 \times 3 \times 3$ Sobel operator or a Canny filter. While Brejl and Sonka [7] used a border appearance model to control the voting process, we use multiple independent filters to identify specific edges. Triangles with weak gradients may be filtered out by defining a threshold $t_{\text{weak edge}}$ for the gradient magnitude that removes the fraction $f_{\text{weak edge}}$ of triangles with the smallest gradient magnitude from R-table construction. In addition, the average gray-value is determined in a $3 \times 3 \times 3$ neighborhood around a triangle center and a desired gray-value range $[g_{\text{edge, min}}, g_{\text{edge, max}}]$ for an edge can be defined. Edges with an average gray-value outside this range are also excluded from R-table construction. Gradient computation as well as gray-value analysis can be done in combination with image sub-sampling.

The normalized image gradient is used to assign the triangle centers of the N_{train} heart meshes to bins of the R-table. For the definition of the bins we use a simple tiling of a spherical surface with rectilinear cells. Here, the size of the cells is specified by means of the angular resolution ϕ . In order to obtain cells with a similar size, we adapt the value of ϕ over the sphere. Furthermore, a merging

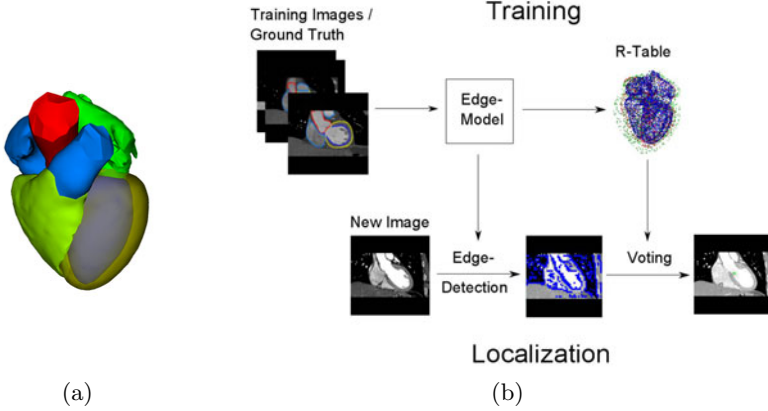


Fig. 2. Heart model (a) and flow chart of our GHT algorithm (b)

step can be applied to reduce the number of R-table entries. For that purpose, an Euclidean grid with edge length d_{merge} is generated and triangle centers of the N_{train} heart meshes in a cell of the grid that belong to the same R-table bin are replaced by a single R-table entry and a weight $w_{R,i}$ defining the frequency of the entry. In a subsequent pruning step, R-table entries can be removed if their weight $w_{R,i}$ is smaller than a threshold $w_{R,\text{Threshold}}$.

For localization, the same pre-processing is applied that has been used for training, i.e. the gradient computation can either be done with a Sobel operator or the Canny filter. Weak edges with a gradient magnitude smaller than the threshold $t_{\text{weak edge}}$ as well as edges with an average gray-value outside the interval $[g_{\text{edge,min}}, g_{\text{edge,max}}]$ can be filtered out. Gradient computation and gray-value analysis can be done in combination with sub-sampling. With regard to the voting and the model point weights, in recent years several extensions of the GHT have been proposed (see e.g. [8] and [9]). For simplicity, we consider two basic variants only, voting may use the weights $w_{R,i}$ or all R-table entries may be attributed the same weight.

2.3 GHT Optimization

Optimization of the GHT parameters and the analysis of design variants require the definition of a suitable metric. In our application the result of the GHT is used to position a (transformed) heart model. In order to quantify how well the this model is positioned in the image, we use the mean constrained point-to-surface distance [2]:

$$\epsilon_{\text{GHT}} = \frac{1}{M} \sum_{i=1}^M \frac{\epsilon_{\Pi(r)}(c_i^{\text{GHT}}, c_i^{\text{ref}}) + \epsilon_{\Pi(r)}(c_i^{\text{ref}}, c_i^{\text{GHT}})}{2},$$

with M as the number of triangles and $\epsilon_{\Pi(r)}(c_i^{\text{GHT}}, c_i^{\text{ref}})$ as the smallest Euclidean distance between the triangle center of the GHT initialized mesh and

a corresponding surface patch in the reference mesh. Here, the surface patch is restricted to triangles within a small geodesic radius ($r = 10.00$ mm) around c_i^{ref} . If the initialization error ϵ_{GHT} is small, the GHT-based model placement is already close to the desired result. Otherwise, the following adaptation steps may fail, and a segmentation failure may be obtained.

To carry out GHT parameter optimization and to analyze design variants, different GHTs were trained and evaluated. In order to access the generalization capabilities of the different variants, and to avoid a biased analysis, 10-fold cross validation was employed. Next to the initialization error ϵ_{GHT} , we recorded the computation time for GHT-based heart localization. Systematic testing of parameter value combinations was done by grid search.

3 Experiments

3.1 Tested Parameter Values

For testing different GHT parameterizations, we used a database with $N_{\text{train}} = 89$ cardiac CTA images with corresponding ground truth annotations in terms of reference meshes. For the generation of the 3D ground truth annotations, automatic segmentations were inspected and refined by a domain expert.

Edge detection was done with the Sobel or the Canny filter. To remove non-specific edges, the threshold $t_{\text{weak edge}}$ was determined to discard the $f_{\text{weak edge}} = 5\%$ weakest edges. In addition, removal of dark edges was tested, i.e. $g_{\text{edge min}}$ was determined in a way that 90% or 95% of the triangles have a larger average gray-value. For image sampling, sub-sampling with 2.0, 3.0, 4.0, and 5.0 mm was considered. An angular resolution of $\phi = 4.0, 6.0, \text{ and } 8.0$ deg was used for the R-table. Merging was offered using an Euclidean edge length matched with image sub-sampling. For pruning, the thresholds $w_{\text{R,Threshold}} = 0.05 \times N_{\text{Train}}$ and $0.10 \times N_{\text{Train}}$ were tested. During localization, three scaling levels (0.8, 1.0, and 1.2) were used. The resolution of the Hough space was matched to the sub-sampling rate for edge detection and gray-value analysis. Voting was tested using the weights $w_{\text{R},i}$ resulting from merging, and identical weights for all R-table entries. In total, this results in 1680 different parametrizations, resp. design variants for the GHT. For a comprehensive evaluation of all combinations, cross validation experiments were carried out using a Windows based compute cluster (5× Dell R710 with 2× Intel(R) Xeon(R) CPU Six Core, X5680 @3.3 GHz, 12 MB cache, 12 GB RAM; Windows HPC Server 2008).

3.2 Parameter Dependence of the GHT

The evaluation of the different parameterizations revealed a broad range of variations with regard to computation time as well as localization accuracy. Fig. 3 shows a subset of the results as a scatter plot, where each dot refers to the average outcome in the 89 CTA images evaluated via 10-fold cross validation for one parametrization. While a large number of different parameterizations resulted in a reasonable localization accuracy, especially the variation in computation time

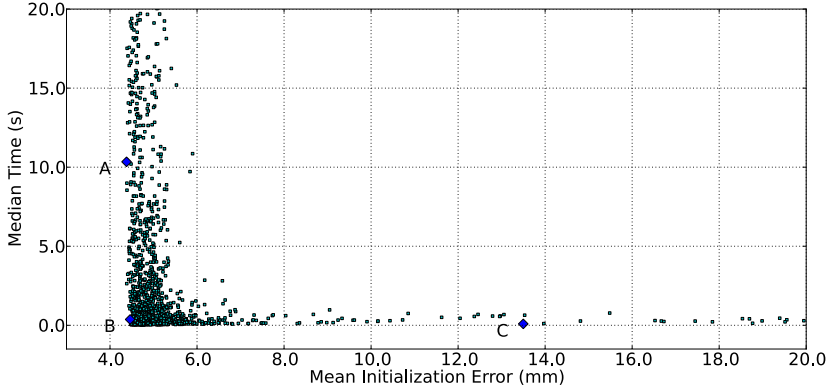


Fig. 3. Overview of the GHT-based localization results. The diagrams show the outcome for different parameterizations in terms of (median) computation time and (average) initialization error ϵ_{GHT} . Three specific configurations have been highlighted (see also section 3.3).

is huge. The scatter plot also shows that there are parameterizations with a good localization accuracy that are very fast at the same time.

Looking in more details at the parameterizations, the results show that a good localization of the heart can be achieved already with coarse image sampling and a coarse resolution of the accumulator (both 5 mm). This is of advantage from a computational perspective as well as for the memory consumption. Merging and pruning did allow to reduce the R-table size and increases the localization speed, while the localization accuracy is only slightly decreased. Voting with identical weights typically outperformed voting with the weights $w_{R,i}$ from merging.

Even though the GHT turned out to be quite robust with respect to the employed edge detection techniques, the best results (in terms of accuracy and speed) were achieved by means of a combination of a Canny filter and gray-value based acceptance intervals. Both techniques remove unspecific edges which resulted in an increased speed and localization accuracy. However, the analysis showed also that an appropriate combination of these techniques is essential. A very restrictive edge detection strategy resulted often in a poor localization accuracy since too many edges were removed from the voting process. This shows that systematic testing is crucial to identify good parameter value combinations.

3.3 Influence of the GHT on Heart Segmentation

Finally, the influence of different GHT parameterizations on the heart segmentation was investigated. This experiment was carried out with 242 CTA datasets that were independent from the data sets used in section 3.1 and 3.2. These datasets represent a broad spectrum of diseases and image quality, and have been acquired with scanners from different vendors at multiple sites. For the evaluation, reference segmentations were obtained by means of a parametric

model adaptation followed by a deformable adaptation. In order to achieve an accurate localization, that is independent from the GHT approach, a manual initialization of the heart was performed by a domain expert using three landmarks (apex, center of mitral valve, center of aortic valve).

Three different parameterizations of the GHT were used in these experiments. They are indicated by A (accurate / slow), B (accurate / fast) and C (inaccurate / fast) in the scatter plot of Fig. 3. These GHT parameterizations were used to initialize the segmentation chain and the resulting segmentation was compared with the reference segmentation. A segmentation was classified as success if the constrained point-to-surface error was < 2 mm and as a failure otherwise. Table 1 summarizes the results. Fig. 4 shows examples of successful segmentations.

Table 1. Influence of the GHT parametrization on the segmentation chain. The initialization error and the median time for localization refer to Fig. 3. Segmentations with a constrained point-to-surface error > 2 mm were classified as failure.

GHT parameterization	Initial. Error	Median Time	CT failures
(A) accurate / slow	4.38 mm	10.3 sec	4 / 242 (1.7%)
(B) accurate / fast	4.46 mm	0.4 sec	6 / 242 (2.5%)
(C) inaccurate / fast	13.50 mm	0.1 sec	90 / 242 (38.0%)

According to Table 1, an improper GHT parametrization can lead to a very poor performance of the entire segmentation chain. With a suitable GHT parametrization, the heart can be detected in a fraction of a second and a failure rate of 2 – 3% of the entire segmentation chain can be achieved. When accepting a computation time in the order of 10 s for heart localization, the failure rate can be slightly reduced.

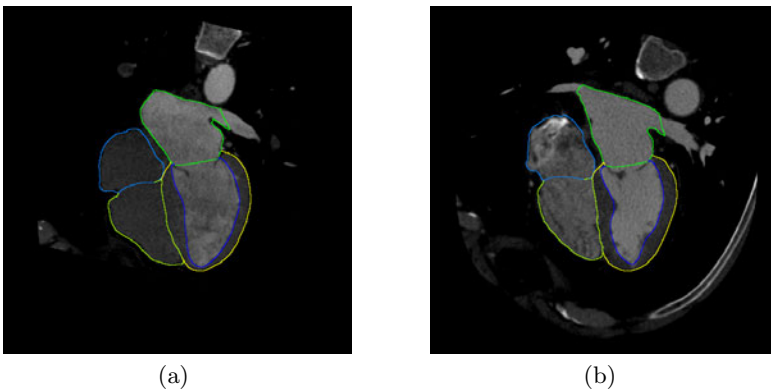


Fig. 4. Example segmentations for two CTA images

4 Conclusions

We presented an approach to optimize the initial organ localization step in a complex segmentation chain consisting of organ localization, parametric organ model adaptation, and deformable model adaptation. In particular, we considered the Generalized Hough Transformation (GHT) for heart localization and heart segmentation in CTA. The GHT performance was measured in terms of the initialization error and computation time. By systematic parameter testing on a computing cluster we identified a GHT parametrization allowing fast and reliable heart localization. A good performance was achieved with coarse image sampling (5 mm), a coarse Hough space resolution (5 mm) and filtering to remove unspecific edges. Finally, we showed that a failure rate for automatic heart segmentation in CTA images of only 2 – 3% can be achieved with a suitable GHT parametrization.

References

1. Hautvast, G., Lobregt, S., Breeuwer, M., Gerritsen, F.: Automatic contour propagation in cine cardiac magnetic resonance images. *IEEE Transactions on Medical Imaging* 25(11), 1472–1482 (2006)
2. Ecabert, O., Peters, J., Schramm, H., Lorenz, C., von Berg, J., Walker, M., Vembar, M., Olszewski, M., Subramanyan, K., Lavi, G., Weese, J.: Automatic Model-Based segmentation of the heart in CT images. *IEEE Transactions on Medical Imaging* 27(9), 1189–1201 (2008)
3. Zheng, Y., Barbu, A., Georgescu, B., Scheuering, M., Comaniciu, D.: Four-Chamber heart modeling and automatic segmentation for 3-D cardiac CT volumes using marginal space learning and steerable features. *IEEE Transactions on Medical Imaging* 27(11), 1668–1681 (2008)
4. Wu, C.F.J., Hamada, M.S.: *Experiments: Planning, Analysis, and Optimization*, 2nd edn. Wiley, Chichester (2009)
5. Peters, J., Ecabert, O., Meyer, C., Kneser, R., Weese, J.: Optimizing boundary detection via simulated search with applications to multi-modal heart segmentation. *Medical Image Analysis* 14(1), 70–84 (2010)
6. Ballard, D.: Generalizing the Hough transform to detect arbitrary shapes. *Pattern Recognition* 13(2), 111–122 (1981)
7. Brejl, M., Sonka, M.: Object localization and border detection criteria design in edge-based image segmentation: Automated learning from examples. *IEEE Transactions on Medical Imaging* 19(10), 973–985 (2000)
8. Maji, S., Malik, J.: Object detection using a max-margin hough transform. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1038–1045 (2009)
9. Leibe, B., Leonardis, A., Schiele, B.: Combined object categorization and segmentation with an implicit shape model. In: *Workshop on Statistical Learning in Computer Vision, European Conference on Computer Vision*, pp. 17–32 (2004)

Automatic Extraction of 3D Dynamic Left Ventricle Model from 2D Rotational Angiocardialogram

Mingqing Chen¹, Yefeng Zheng¹, Kerstin Mueller^{2,3}, Christopher Rohkohl²,
Guenter Lauritsch², Jan Boese², Gareth Funka-Lea¹,
Joachim Hornegger³, and Dorin Comaniciu¹

¹ Image Analytics and Informatics, Siemens Corporate Research, Princeton, NJ, USA

² Healthcare Sector, Siemens AG, Forchheim, Germany

³ Pattern Recognition Lab, University Erlangen-Nuremberg, Germany
yefeng.zheng@siemens.com

Abstract. In this paper, we propose an automatic method to directly extract 3D dynamic left ventricle (LV) model from sparse 2D rotational angiocardialogram (each cardiac phase contains only five projections). The extracted dynamic model provides quantitative cardiac function for analysis. The overlay of the model onto 2D real-time fluoroscopic images provides valuable visual guidance during cardiac intervention. Though containing severe cardiac motion artifacts, an ungated CT reconstruction is used in our approach to extract a rough static LV model. The initialized LV model is projected onto each 2D projection image. The silhouette of the projected mesh is deformed to match the boundary of LV blood pool. The deformation vectors of the silhouette are back-projected to 3D space and used as anchor points for thin plate spline (TPS) interpolation of other mesh points. The proposed method is validated on 12 synthesized datasets. The extracted 3D LV meshes match the ground truth quite well with a mean point-to-mesh error of $0.51 \pm 0.11mm$. The preliminary experiments on two real datasets (included a patient and a pig) show promising results too.

1 Introduction

The real time angiocardialogram on a C-arm system is the workhorse imaging modality for many cardiac interventions. Nowadays it is becoming desirable to create a dynamic 3D left ventricle (LV) model automatically, which could be served to 1) quantitatively evaluate cardiac function, such as LV volume, LV ejection fraction (EF) and regional wall motion during the intervention, 2) provide visual guidance during intervention by overlaying this model onto 2D fluoroscopy, and 3) reduce motion artifacts for 3D image reconstruction [1]. A direct approach is to segment the LV from cine CT volumes, which are derived by electrocardiogram (ECG) gated reconstruction [2]. However, the X-ray source/detector needs to sweep 5-6 times to capture enough 2D projection data for each cardiac phase. The patient is required to hold breath during the whole

procedure of 30 s (which is difficult to a very sick patient) to remove the respiratory motion. Furthermore, longer acquisition time incurs more radiation dose, which is a big concern nowadays.

Alternatively, one can generate a dynamic mesh directly from 2D projections. There are some previous studies using biplane and multi-view systems to recover thorax and lung [3,4] and the LV surface [5,6,7]. In Lotjonen et al.'s work, a prior model, which is a representative of average anatomy, is projected onto each 2D angle. The virtual projection is matched with input image using free form deformation [3,4]. Veistera et al. used triangulated mesh of the LV segmented from an MR image as prior model and projected the mesh onto each 2D projection. Manually annotated contour is used to deform the mesh. Moriyama et al. [6] proposed an iterative framework to recover LV mesh from multi views by fitting a 4D surface model based on B-splines. However, the geometry is limited to a star shape due to the spherical coordinates used.

In this paper, we propose a fully automatic method to fit a dynamic 3D LV model using projection images generated by a single rotation of C-arm. The proposed method has several advantages: 1) Since the prior LV model is segmented from an ungated volume reconstructed from the 2D projections, the 3D to 2D mapping is straightforward. 2) The detection of LV blood pool boundary is efficient. The searching is confined to the profiles of silhouette points of projected mesh, without filtering the entire image. 3) Thin plate spline (TPS) interpolation is used to compute the deformation of the whole mesh. The inclusion of smoothing factor in TPS allows correction of outliers in estimated deformation vectors.

2 Method

The main steps of our approach are illustrated in Fig. 1 with the intermediate results shown in Fig. 2. Firstly a 3D CT volume is reconstructed using all 2D projections. Though with significant cardiac motion artifacts, the image quality is good enough for automatic extraction of a rough 3D LV endocardium mesh model using the approach proposed in [8] (Fig. 2a). We then project this static mesh onto each 2D image (Fig. 2b). The silhouette is extracted from the projected mesh and adjusted along the normal direction to the optimal position

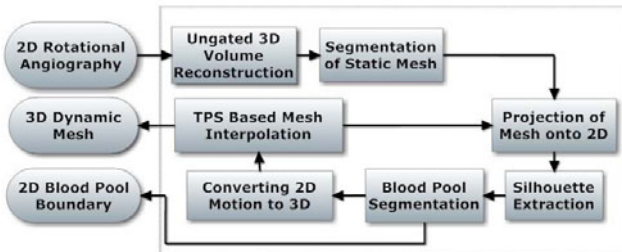


Fig. 1. Flowchart of the extraction of 3D dynamic left ventricle meshes

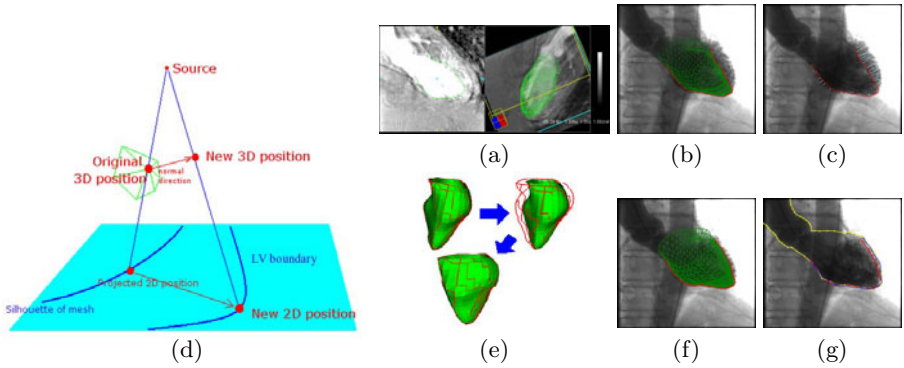


Fig. 2. Intermediate results of the proposed method. (a) Ungated 3D reconstruction volume overlaid with a static left ventricle (LV) mesh. (b) Projected static mesh (green) and its silhouette (red). (c) Deforming the silhouette to fit the LV blood pool boundary (blue). (d) Converting 2D deformation to 3D. (e) Warping the whole mesh using the silhouettes as anchor points. (f) Re-projection of deformed mesh. (g) Comparison of the mesh silhouette (red), segmented (blue) and annotated (yellow) LV blood pool.

using a learning based boundary detector (Fig. 2e). The 2D deformation vector is then back-projected to 3D (Fig. 2f). The patient image we are working on has 133 projections, which is acquired in 5 s. For a typical heart rate of 70 bpm, each cardiac phase has only five projections. Thus the mesh deformation is determined by points from five silhouettes, which is used as anchor points for TPS interpolation (Fig. 2e). In this way a sequence of 3D meshes are generated. The estimated dynamic mesh can further be re-projected to corresponding 2D angiographic images for further refinement again. The iterative procedure converges in a few iterations.

2.1 Mesh Silhouette Extraction

On a 2D projection image, the LV blood pool filled with contrast agent is the most salient image boundary. The LV mesh is deformed to make its silhouette match the blood pool boundary. An efficient approach is proposed to extract the silhouette of a projected mesh based on the connectivity of mesh edges. It should be noted that the silhouette is not only composed with full length of original mesh edges. Two edges not coplanar in 3D may intersect in a 2D projection and only part of each is on the silhouette (Fig. 3). The general procedure is as follows. 1) Select candidate mesh edges for the silhouette; 2) Split the candidate edges at the intersection points; and 3) Edge following to extract silhouette. The efficiency of our approach comes from two facts: 1) The first step is accomplished by topological analysis of the projected mesh. For a triangulated mesh (e.g., our heart model), an edge cannot be on the silhouette if it is shared by two triangles on the opposite sides of this edge. Thus most edges are excluded for further investigation in this step. 2) The second step determines whether an intersection point exists or not for every remaining edge pairs. The computation of

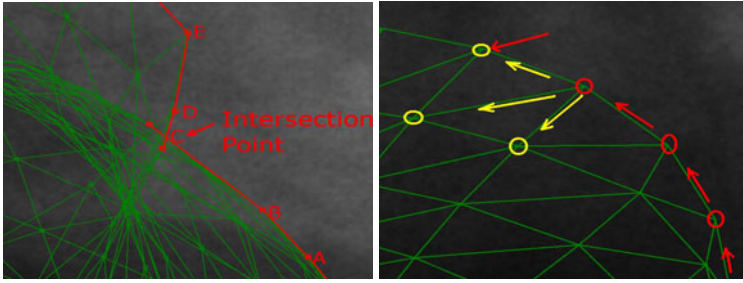


Fig. 3. Mesh silhouette extraction. **Left:** Splitting candidate silhouette edges (red) at the intersection (point C). **Right:** Edge following, which consistently follows the right-most edge.

intersection is time consuming. It is sped up by performing a quick condition check based on bounding box, which uses light operations to reject majority of the non-intersecting edge pairs. Thus the time consuming intersection computation is applied to only a few edge pairs.

Finally, an edge following procedure is taken to extract the silhouette from the split edges (Fig. 3). It is initialized with the right-most point of the projected mesh, which is guaranteed to be on the silhouette. Initialized with the upward search direction, it iteratively follows the right-most edge until coming back to the starting point. The whole path composes the extracted mesh silhouette.

2.2 Blood Pool Segmentation

During the training stage, the manually annotated LV blood pool boundary is given as input. For each boundary point, the steerable features [8] of pixels along the normal direction of this point are taken to extracted for a probabilistic boosting tree (PBT) classifier, which is a combination of the decision tree and AdaBoost classifier.

Let's assume the silhouette is composed of N points. During the segmentation stage, for each silhouette point n , a profile with predefined length K along the normal direction is extracted. A probability value is computed for each pixel on the profile based on PBT. Thus an $N \times K$ 2D array $A(n, k_n)$ can be computed, where n denotes the silhouette point index and k_n represents the pixel index along profile n . It is desirable to deform the silhouette contour to pixels with maximized summation of probability value, under some smoothness constraint that the difference of displacements between neighboring points is restricted to be smaller than λ . Mathematically the optimization problem has the form as:

$$\max \sum_{n=1}^N A(n, k_n), \text{ under the constraint } \begin{cases} |k_1 - k_N| < \lambda \\ |k_{n+1} - k_n| < \lambda, \end{cases} \quad (1)$$

It can be solved with an optimal graph search algorithm [9], where a 1D loop surface can be detected in a 2D graph.

2.3 Back-Projection of Deformation Vectors

To estimate the 3D motion, we assume that each mesh point is moving along its surface normal. It is a reasonable assumption since the LV motion is dominated by the contraction and dilation of the chamber. For each silhouette point \mathbf{M} , we project its 3D normal vector \mathbf{n} onto 2D to compute the 2D normal vector \mathbf{n}_P

$$\mathbf{n}_P = \frac{(\mathbf{P}(\mathbf{M}^+ + \mathbf{n}^+) - \mathbf{P}\mathbf{M}^+)^-}{\|(\mathbf{P}(\mathbf{M}^+ + \mathbf{n}^+) - \mathbf{P}\mathbf{M}^+)^-\|}, \quad (2)$$

where \mathbf{P} is a 4×3 projection matrix for 3D to 2D mapping. Superscript $^+$ denotes representation of original vectors using the homogeneous coordinate by adding “1” as another dimension of perspective plane, while superscript $^-$ represents using the original vector (For 2D, (x, y, s) is converted as $(x/s, y/s)$). The distance of 3D displacement (t) can be calculated by solving the following equation,

$$\mathbf{P}(\mathbf{M}^+ + (\mathbf{n}t)^+) = \mathbf{M}'_P, \quad (3)$$

where \mathbf{M}'_P is the 2D homogeneous coordinate of blood pool boundary, which is also the projection of the mesh point’s new location \mathbf{M}' . By expanding Eqn. (3), we get

$$\frac{P_{11}(M_x + n_x t) + P_{12}(M_y + n_y t) + P_{13}(M_z + n_z t) + P_{14}M_w}{(P_{31}(M_x + n_x t) + P_{32}(M_y + n_y t) + P_{33}(M_z + n_z t) + P_{34}M_w)} = M'_{Px}, \quad (4)$$

where P_{ij} the element at the i^{th} row and j^{th} column of projection matrix \mathbf{P} . (M_x, M_y, M_z, M_w) is the coordinate of \mathbf{M}^+ , and M'_{Px} is the x coordinate of the 2D blood pool boundary point.

2.4 TPS Interpolation

The 3D positions of non-silhouette mesh points need to be interpolated. The thin plate spline (TPS) [10] is a popular coordinate interpolation approach by minimizing a physical energy function,

$$E = \sum_{i=1}^k \|f(x_i) - y_i\|^2 + \lambda \iint_{R^2} \left[\left(\frac{\partial^2 z}{\partial x^2} \right)^2 + 2 \left(\frac{\partial^2 z}{\partial x \partial y} \right)^2 + \left(\frac{\partial^2 z}{\partial y^2} \right)^2 \right] dx dy, \quad (5)$$

where the first term is the interpolation error of the anchor points and the second term is the bending energy of the transformation of a “thin plate.” The parameter λ is tuned appropriately to control the balance between exact mapping of anchor points and rigidity of deformation.

3 Experiments

We first validate the proposed method on synthesized datasets using the 3D LV meshes extracted from 12 4D CT scans, on which we know both the 2D and 3D ground truth. The whole system is also tested on two real datasets, including a patient and a pig.

3.1 Validation on Synthesized Data

In this experiment, we validate the basic assumptions (e.g., the 3D motion direction along the surface normal and TPS interpolation) used in the proposed method without considering the blood pool segmentation error. The dynamic LV mesh sequence extracted from 12 4D CT scans are used as ground truth. The imaging geometry of one real patient data is used, which has 133 projections. Originally the 4D CT has 10 phases. The dynamic meshes are resampled to 26 phases to mimic the five cardiac cycles during image acquisition. The averaged mesh of the dynamic sequence is computed as the prior model. The proposed method is applied to estimate the dynamic sequence, which is compared to the real sequence to measure the accuracy.

LV volumes and point-to-mesh error are used to measure the difference between computed meshes and ground truth. Fig. 4 shows the convergence of LV volume error on 12 sequences through five iterations of the algorithm, where the error is reduced quickly in the first two iterations. The initial volume error is about 30 ml. We achieve a mean volume error of 3.27 ± 0.71 ml and relative volume error of $4.0 \pm 4.4\%$. As shown by the volume curve in Fig. 4, the estimated LV volume matches the ground truth quite well. The estimated ejection fraction (EF) of 43.7% is close to the true value of 45.8%. For the point-to-mesh error, the initial mean error is 2.19 ± 0.56 mm, and the final error is significantly reduced to 0.51 ± 0.11 mm, which is smaller than the mesh resolution (the mean mesh edge length is 4.29 mm).

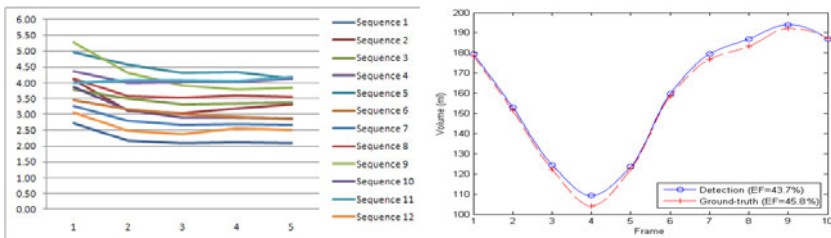


Fig. 4. Validation on synthesized data. **Left:** LV volume error (ml) of 12 mesh sequences plotted against the number of iterations; **Right:** The estimated LV volume curve and ejection fraction (EF) against the ground truth for one sequence.

3.2 Validation on Real Data

The proposed method is also validated on two real datasets, including a patient and a pig. For the patient dataset, about five cardiac cycles were captured in the image sequence with 133 projection images. The frame rate of the pig dataset was set much higher, resulting in 395 2D projection images, spanning 13 cardiac cycles. The ECG signal is recorded to assign a cardiac phase to each 2D projection. Since we do not have ground truth of the 3D mesh, we measure the distance of the mesh silhouette to the annotated blood pool boundary. The error of the automatically segmented blood pool (using the approach presented in Section 2.2) is also validated, with the mean error of 2.35 ± 1.32 mm on the patient

data. The error of the re-projected mesh silhouette is slightly higher, 2.84 ± 1.44 mm. The 2D angiographic images are very noisy and the LV could be occluded by other highly contrasted structures, e.g., the aorta and the diaphragm. To study the inter-observer variability, two clinicians were asked to annotate the blood pool independently. The mean error between the two manual annotations is 1.60 ± 0.87 mm. Fig. 5 shows the extracted LV mesh and the projected silhouette on two frames. Our approach is also tested on a pig dataset, where the image contrast is much lower and there are strong confounding boundaries nearby such as the LV epicardium. The mesh silhouette error after four iterations is 5.60 ± 2.05 mm.

Using the estimated 3D cardiac motion, we perform motion-compensated reconstruction [11]. As shown in Fig. 6 on the patient dataset, the trabeculations of the LV is clearly visible after motion compensated.

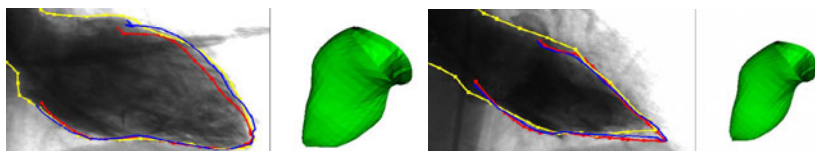


Fig. 5. Estimated 3D mesh (green mesh) and the re-projected mesh silhouette (red contour) on two angiographic images. For comparison, the automatically segmented blood pool boundary (blue) and manual annotation (yellow) are also shown.

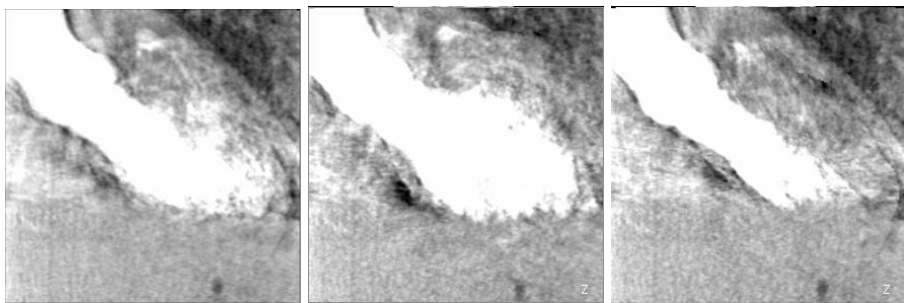


Fig. 6. Reconstructed CT volumes using the estimated 3D motion for the patient dataset. **Left:** Ungated reconstruction. **Middle and Right:** Motion-compensated reconstruction for the end-diastolic and end-systolic phases, respectively.

4 Conclusion

We proposed a novel method to automatically extract the 3D dynamic LV model from 2D rotational angiographic images captured in a single rotation of a C-arm system. The estimated dynamic model can be used to extract several important measurements (e.g., volume and EF) of the LV function and it can also

be overlaid onto 2D fluoroscopic images to provide visual guidance during cardiac interventions. The experiment on synthesized data validated the appropriateness of the proposed method. The feasibility study on two real datasets (a patient and a pig) showed promising results and we plan to test on more real datasets in the future. Our approach can be extended to extract other organs with cyclic motion (e.g., other heart chambers and the aorta).

Acknowledgments. We would like to thank to Drs. Patrick W. Serruys, Carl Schultz, Peter de Jaegere, and Robert van Geuns from Thorax Center, Erasmus MC, Rotterdam, The Netherlands for acquiring data.

References

1. Prummer, M., Hornegger, J., Lauritsch, G., Wigstrom, L., Girard-Hughes, E., Fahrig, R.: Cardiac C-arm CT: A unified framework for motion estimation and dynamic CT. *IEEE Trans. Medical Imaging* 28(11), 1836–1849 (2009)
2. Lauritsch, G., Boese, J., Wigstrom, L., Kemeth, H., Fahrig, R.: Towards cardiac C-arm computed tomography. *IEEE Trans. Medical Imaging* 25, 922–934 (2006)
3. Lotjonen, J., Magnin, I., Reinhardt, L., Nenonen, J., Katila, T.: Automatic reconstruction of 3D geometry using projections and a geometric prior model. In: Taylor, C., Colchester, A. (eds.) *MICCAI 1999*. LNCS, vol. 1679, pp. 192–201. Springer, Heidelberg (1999)
4. Lotjonen, J., Magnin, I., Nenonen, J., Katila, T.: Reconstruction of 3-D geometry using 2-D profiles and a geometric prior model. *IEEE Trans. Medical Imaging* 18(10), 992–1002 (1999)
5. Veistera, H., Lotjonen, J.: Reconstructing 3D boundary element heart models from 2D biplane fluoroscopy. *Proc. Functional Imaging and Modeling of the Heart*, 17–23 (2001)
6. Moriyama, M., Sato, Y., Naito, H., Hanayama, M., Ueguchi, T., Harada, T., Yoshimoto, F., Tamura, S.: Reconstruction of time-varying 3D left ventricular shape from multiview x-ray cineangiograms. *IEEE Trans. Medical Imaging*, 773–785 (2002)
7. Chhatiwalla, A., Kramer, C., Peixoto, A., Samady, H.: Measurement of left ventricular mass by contrast ventriculography. *Texas Heart Inst. J.* 31(7), 323–327 (2008)
8. Zheng, Y., Barbu, A., Georgescu, B., Scheuring, M., Comaniciu, D.: Four-chamber heart modeling and automatic segmentation for 3D cardiac CT volumes using marginal space learning and steerable features. *IEEE Trans. Medical Imaging* 27(11), 1668–1681 (2008)
9. Li, K., Wu, X., Chen, D.Z., Sonka, M.: Optimal surface segmentation in volumetric images—A graph-theoretic approach. *IEEE Trans. Pattern Anal. Machine Intell.* 28, 119–134 (2006)
10. Bookstein, F.: Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Trans. Pattern Anal. Machine Intell.* 11(6), 567–585 (1989)
11. Schäfer, D., Borgert, J., Rasche, V., Grass, M.: Motion-compensated and gated cone beam filtered back-projection for 3-D rotational X-ray angiography. *IEEE Trans. Medical Imaging* 25(7), 898–906 (2006)

Automatic View Planning for Cardiac MRI Acquisition

Xiaoguang Lu^{1,*}, Marie-Pierre Jolly¹, Bogdan Georgescu¹, Carmel Hayes²,
Peter Speier², Michaela Schmidt², Xiaoming Bi³, Randall Kroeker⁴,
Dorin Comaniciu¹, Peter Kellman⁵, Edgar Mueller², and Jens Guehring²

¹ Image Analytics and Informatics, Siemens Corporate Research, Princeton NJ, USA
xiaoguang.lu@siemens.com

² Siemens AG, Healthcare Sector, H IM MR PLM-AW CARD, Erlangen, Germany

³ Siemens Medical Solutions USA, Chicago IL, USA

⁴ Siemens Medical Solutions Canada, Winnipeg MB, Canada

⁵ National Institutes of Health, Bethesda MD, USA

Abstract. Conventional cardiac MRI acquisition involves a multi-step approach, requiring a few double-oblique localizers in order to locate the heart and prescribe long- and short-axis views of the heart. This approach is operator-dependent and time-consuming. We propose a new approach to automating and accelerating the acquisition process to improve the clinical workflow. We capture a highly accelerated static 3D full-chest volume through parallel imaging within one breath-hold. The left ventricle is localized and segmented, including left ventricle outflow tract. A number of cardiac landmarks are then detected to anchor the cardiac chambers and calculate standard 2-, 3-, and 4-chamber long-axis views along with a short-axis stack. Learning-based algorithms are applied to anatomy segmentation and anchor detection. The proposed algorithm is evaluated on 173 localizer acquisitions. The entire view planning is fully automatic and takes less than 10 seconds in our experiments.

1 Introduction

Accurate morphological and functional measurements of the heart are essential in clinical applications for diagnosis, prognostic, and therapeutic decisions. Magnetic resonance imaging (MRI) allows morphological characterization of heart structures with precision. For the past decade, there has been a great deal of efforts focused on automatic cardiac segmentation [5][12], however, less attention has been paid to automatic and fast planning for cardiac MRI acquisition [7][8][4][6], which still remains challenging in clinical practice.

During conventional cardiac MRI acquisition, anchoring the heart is typically performed using a multi-step approach involving the acquisition of double-oblique slices in order to localize the long and short-axis of the heart. Based on those localizers the standard heart views (e.g, 2-chamber, 3-chamber, 4-chamber

* Corresponding author.

and short-axis views from base to apex) are planned. This approach is operator-dependent and time consuming. For example, a typical protocol starts with capturing coronal localizers, followed by multi-slice localizer at different orientations in order to align the left ventricle (LV) with the isocenter of the scanner; then one vertical long-axis localizer is planned on transversal slice and one horizontal long-axis localizer is planned on vertical long-axis view, from which a few short-axis views are planned; based on the short-axis views, 2-, 3-, and 4-chamber views are planned; and finally the short axis stack is planned on the 4-chamber view. It requires detailed knowledge of the heart for operators to plan the views at every step during this long process, while the patient is in the scanner.

As parallel MR imaging techniques rapidly advance, single breath-hold acquisitions with full chest coverage become feasible. Instead of acquiring a series of 2D localizer slices with operator interventions, a direct approach to prescribing the clinically relevant slices in a fully automated fashion from a single acquisition would be highly beneficial. However, this is a challenging task in order to be clinically viable because it requires a robust approach to differentiating the heart from other complex anatomical structures in a full-chest volume, accounting for large variations across populations, and with quick response time.

Lelieveldt et al. [7,8] proposed fuzzy implicit surface templates to model thoracic anatomies and automatically estimate the LV orientation for view planning. The method was evaluated on 20 volumes and planning was for short-axis stack only. Danilouchkine et al. [4] quantitatively evaluated the difference of clinical measurements derived from acquisitions between manual and automatic planning on 10 healthy volunteers, concluding that automated CMR planning methods could provide accurate measurements of LV dimensions in normal subjects. In addition to the short-axis stack, long-axis images also provides clinical values, such as through-plane motion analysis [2]. Jackson et al. [6] proposed a semi-automatic approach for planning short-axis stack, and two long-axis views, i.e., HLA (4-chamber view) and VLA (2-chamber view). User interaction is needed for chamber segmentations, which were based on an EM algorithm. The approach was evaluated on 12 volunteers.

We propose a fully automatic algorithm to prescribe short-axis stack and standard long-axis view, including 2-, 3-, and 4-chamber views, from a single 3D MR volume for cardiac MRI acquisition. Our approach to improving the clinical workflow is based on machine learning based methods to localize and delineate cardiac anatomies in a 3D volume, and detect a set of cardiac landmarks to anchor chambers in order to prescribe the views. The proposed approach is evaluated on 173 localizer volumes from 100 subjects.

2 Methodology

2.1 Overall Workflow

Our view planning workflow for cardiac MRI acquisition as shown in Fig. 1 contains the following steps: 1) 3D static volume acquisition; 2) LV localization (pose

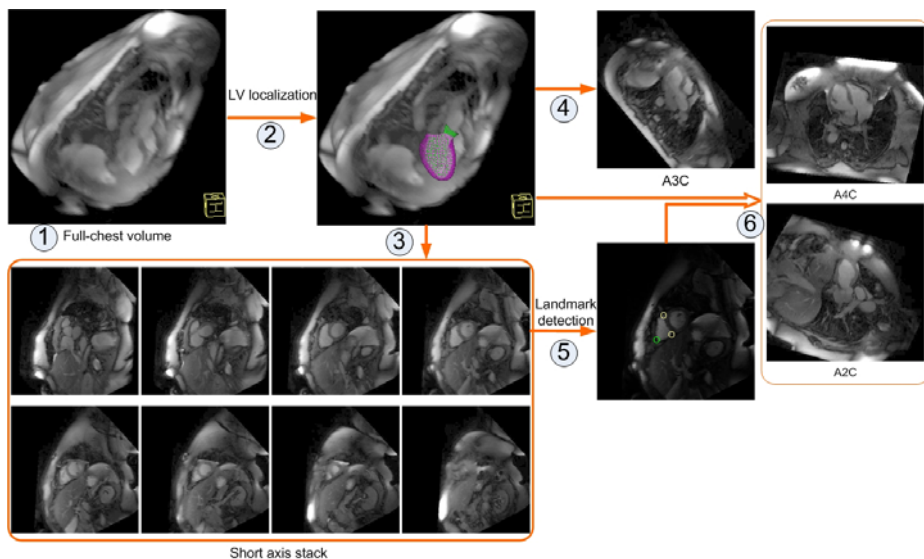


Fig. 1. System workflow, Numbers indicate the steps illustrated in Sec. 2.1

estimation) and delineation (boundary extraction); 3) short axis stack prescription; 4) 3-chamber view calculation; 5) landmark detection in mid-ventricular slice(s); 6) 4-chamber and 2-chamber view calculations.

For LV delineation and landmark detection, we adapted the basic approaches presented in [12] and [9] due to computational efficiency and accuracy. For completeness, we briefly describe the methods in our application context.

A mesh representation of the anatomical shape is created using a database of 3D volumes that are manually delineated. For left ventricle, this anatomically accurate model is composed of LV endocardium, LV epicardium, and LVOT. The LV model is a triangular mesh as shown Fig. 2. It is used to fit to a given 3D cardiac volume to delineate corresponding anatomical structures. Including LVOT into our workflow helps anchor the 3-chamber view.

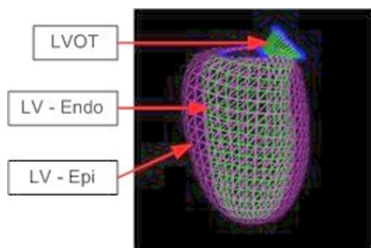


Fig. 2. Triangular mesh model for left ventricle and LV outflow tract (LVOT)

2.2 LV Localization and Delineation

The LV pose (i.e., positions, orientations, and scales) in a 3D volume is first estimated at the LV localization stage. At the subsequent delineation stage, local deformations of the model with the estimated pose are applied in order to fit the model to LV boundaries.

Learning-based object detection approaches have been demonstrated successful in many applications [11]. Learning-based methods trained on expert annotations are appropriate to handle complex appearance and heterogeneous characteristics of anatomical features in medical images, as the complex prior knowledge is implicitly encoded.

In order to estimate the LV pose, we need to solve for the nine-parameter space, including three translations, three orientations, and three scales. We design a series of detectors that estimate parameters at a number of sequential stages in the order of complexity, i.e., translation, orientation, and scale. Multiple hypotheses are maintained between stages, which quickly removes false hypotheses at the earlier stages while propagating the right hypothesis to the final stage. Only one hypothesis is selected as the final detection result.

A large database is collected with LV annotated. For each detector, a set of positive and negative samples are extracted from the annotated database with positive samples as the true targets (close to the ground truth) and negative samples as others (far away from the ground truth). The detection task is a binary classification. Our detectors to estimate the LV pose and extract LV boundaries are constructed using probabilistic boosting trees (PBT) [10], which select a set of discriminative features that are used to distinguish the positive samples from negatives from a large pool of features. The classifier is a tree-based structure with which the posterior probabilities of the presence of the object of interest are calculated from given data. The nodes in the tree are constructed by a nonlinear combination of simple classifiers using boosting techniques. For the detector at the translation stage, we choose 3D Haar wavelet-like features, which are calculated efficiently using integral image-based techniques. For the detectors at the orientation and scale search stages, steerable features [12] are applied, whose computation does not require volume rotation and re-scaling, which are computationally expensive, especially when the hypothesis search space is large. The boundary detectors also use steerable features.

With the model pose estimated, we align the mean shape (average model of all annotations) with data to get an initial estimate of the object shape. To capture the true anatomical morphology of the LV shape, we deform the mean LV model by searching the boundary for each vertex of the model. The boundary hypotheses are taken along the normal directions at each vertex of the mean model. Detection is achieved using a boundary detector trained by PBT with steerable features. The detected boundaries are constrained by projecting the detected model onto a shape subspace obtained by the annotated dataset, which was constructed using principal component analysis [3]. As a result, the mesh model is fitted to given volumetric data in alignment with the LV.

2.3 View Planning

With the left ventricle delineated, a few cardiac anchors are computed. In our mesh model, the LV base and apex have corresponding vertices. The LV base center is calculated as the centroid of the vertices on the LV base. The LV long axis is calculated by connecting the LV base center and the apex, which is also used to plan the short-axis stack. The method of fitting a line across centers of all short axis MPRs as shown in [7] and [6] is also applicable. The entire short axis stack can then be planned from base to apex. To plan the 2- and 4- chamber views, we reconstruct a mid-ventricular short axis slice (MPR) from the short axis stack, and apply the landmark detection algorithm proposed in [9] to localize the anterior and inferior RV insertion points. The 2-chamber view is calculated to be in parallel to the line connecting the two RV insertion points and cross the LV blood pool center that lies on the short-axis MPR used for landmark detection, which is obtained through the detected LV mesh. The 4-chamber view is computed to cross the LV blood pool center and the RV lateral point (the green landmark in Fig. 1). Fig. 1 presents an example of all planned views.

3 Experiments

A 3D full-chest MR scan was obtained through parallel imaging within a single breath-hold. A single volume was acquired at mid-diastole using an ECG gated segmented acquisition with T2-prepared SSFP readout with chemical shift fat suppression. Typical protocol parameters are: $400 \times 400 \times 220 \text{ mm}^3$ FOV prescribed as a coronal slab with matrix of $256 \times 202 \times 44$, corresponding to $1.6 \times 2 \times 5 \text{ mm}^3$ resolution, interpolated to 2.5 mm slices. Images are acquired on 1.5T Siemens MAGNETOM Avanto/Espree with a 32 channel coil enabling parallel imaging with acceleration rate $R=3 \times 2$ (PE in LR direction \times PAR in AP direction) and $3/4$ partial Fourier in PAR dimension. Breath-hold duration is typically less than 20s with all PE lines acquired in a single shot per heartbeat. 88 volumes from 77 patients were acquired under this protocol, which we called the baseline protocol. In addition, a few experimental localizer protocols were applied to investigate the image variations by modifying acquisition parameters such as the number of breathholds, the number of slices per heart beat, real-time acquisition, etc., resulting in additional 85 volumes from a combination of 23 volunteers and patients. In total, 173 localizer volumes from 100 subjects were collected. Examples acquired with different protocols are provided in Fig. 3.

For each volume, the left ventricle was manually delineated using triangular mesh representations as ground truth, including endocardium, epicardium, and LVOT boundaries. The standard 2-, 3-, 4-chamber views were also manually identified by visually navigating the volume in 3D, with each view recorded by a combination of a plane normal vector and an anchor on the plane. The LV long axis groundtruth is calculated as a vector from the LV apex pointing toward LV base center (the centroid of mesh points in the mitral plane), which was used as the normal direction of the short-axis stack.

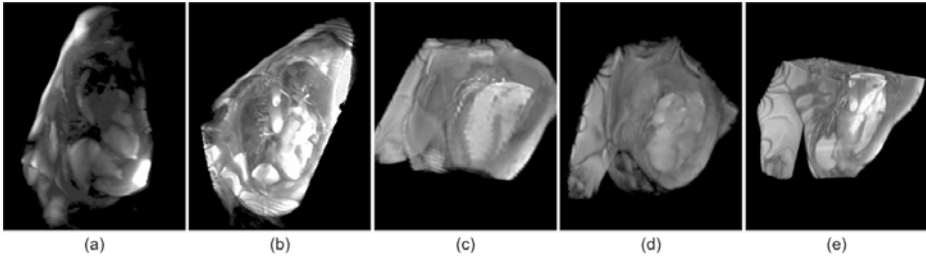


Fig. 3. Examples of volumes captured with different protocols in our database. (a)-(b) Acquisitions using the baseline protocol. (d)-(e) Acquisitions with the variants of the baseline protocol.

A 4-fold cross-validation scheme was applied for evaluation. The entire 173 volumes were randomly partitioned into four quarters. No volumes from the same subject were partitioned into different quarters. For each fold evaluation, three quarters were combined for training and the remaining one quarter was used as unseen data for testing. This procedure was repeated four times so that each volume has been used once for testing. To evaluate the LV delineation results, for each fitted mesh, the distance from each vertex to the groundtruth mesh (manual annotation) was computed (so called point-to-mesh distance). The average of distances from all vertices of the fitted mesh was used as the measurement. The mean, standard deviation, and median values of this distance measurement are 4.96mm, 10.35mm, and 2.51mm, respectively, which indicate a good delineation and anchoring for the subsequent view calculation. Large standard deviations are due to detection outliers.

Fully automated view planning examples are provided in Fig. 4. For short-axis stack planning evaluation: the LV long axis (i.e., the normal of the prescribed short-axis stack) is used to evaluate short-axis stack axis direction estimation accuracy, which is calculated as the angle between the estimated LV long axis and the annotated ground truth; to evaluate the coverage of the prescribed short-axis stack, the groundtruth LV base and LV apex are projected onto the estimated LV long axis, then the distances from the projected groundtruth to the automatically estimated base and apex positions are computed. For long-axis view planning evaluation: each estimated view is represented by its normal vector and an anchor on the plane; the angle difference is calculated between the normal vector and the ground truth normal direction; the anchor position difference from the groundtruth is computed as the point-to-plane distance from the anchor on the detected view to the annotated groundtruth view plane; the anchor on the detected view is the LV blood pool center on the prescribed mid-ventricle slice (see Sec. 2.3) for landmark detection. Performance is summarized in Table 1. Large errors are mostly due to image artifact and poor image quality such as the examples shown in Figs. 3(c) and (d).

On average, it took about 5 seconds to localize and delineate the LV from a single full-chest volume (e.g, $256 \times 256 \times 112$), and about half a second to detect landmarks from a mid-ventricular slice. The entire LV anchoring and view planning process took about 7 seconds on an Intel core-duo 2.66GHz CPU.

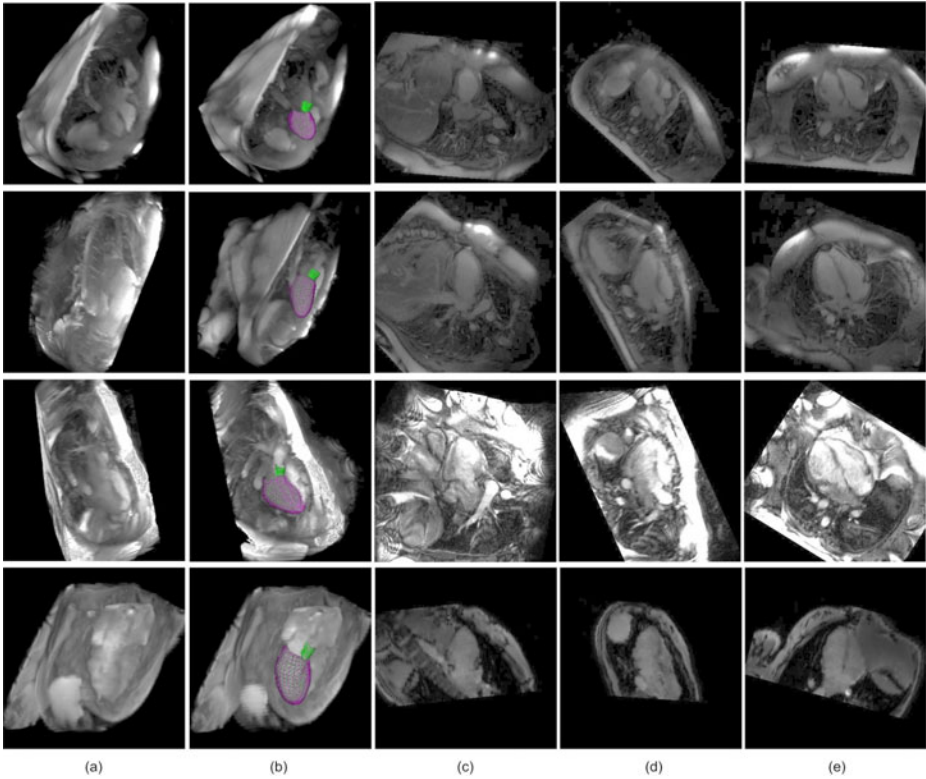


Fig. 4. Automatic view planning examples. (a) Original 3D volume. (b) LV delineation results. The prescribed 2-, 3-, and 4-chamber views are provided in (c)-(e), respectively.

Table 1. Automatic view planning evaluation by a 4-fold cross validation. Axis and angle differences are measured in *degrees* and position difference is measured in *mm*.

	Short-axis stack			2-chamber view		3-chamber view		4-chamber view	
	Axis	Base	Apex	Angle	Position	Angle	Position	Angle	Position
Mean	8.6	13.3	11.7	18.9	6.6	12.3	4.6	17.6	5.7
Std	9.7	16.7	20.6	21.0	8.8	11.0	7.7	19.2	8.5
Median	5.5	7.4	6.1	10.7	4.0	8.9	2.6	9.8	3.7

4 Conclusions

We have presented a fully automatic and fast approach to view planning for cardiac MRI acquisition. Even in the presence of large image variations due to a mixture of experimental protocols, the proposed approach shows its capability to handle a majority of cases with reasonable image quality. With this approach, time for LV acquisition planning can be significantly reduced. Our future focus is to further improve view planning accuracies to overcome poor image quality and large image variation challenges.

References

1. Cardiac MR Left Ventricle Segmentation Challenge. In: MICCAI Workshop (2009), http://smial.sri.utoronto.ca/LV_Challenge/Home.html
2. Attili, A.K., Schuster, A., Nagel, E., Reiber, J.H.C., van der Geest, R.J.: Quantification in cardiac MRI: advances in image acquisition and processing. *Int. J. Cardiovasc Imaging* 26, 27–40 (2010)
3. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active shape models - their training and application. *Comp. Vision and Image Understanding* 61, 38–59 (1995)
4. Danilouchkine, M.G., Westenberg, J.J.M., Reiber, J.H.C., Lelieveldt, B.P.F.: Accuracy of short-axis cardiac MRI automatically derived from scout acquisitions in free-breathing and breath-holding modes. *Magnetic Resonance Materials in Physics, Biology and Medicine* 18(1), 7–18 (2005)
5. Frangi, A., Niessen, W., Viergever, M.: Three-dimensional modeling for functional analysis of cardiac images: A review. *IEEE Trans. on Medical Imaging* 20(1) (2001)
6. Jackson, C.E., Robson, M.D., Francis, J.M., Noble, J.A.: Computerised planning of the acquisition of cardiac MR images. *Computerized Medical Imaging and Graphics* 28(7), 411–418 (2004)
7. Lelieveldt, B.P.F., van der Geest, R.J., Lamb, H.J., Kayser, H.W.M., Reiber, J.H.C.: Automated observer-independent acquisition of cardiac short-axis MR images: A pilot study. *Radiology* 221, 537–542 (2001)
8. Lelieveldt, B.P.F., Sonka, M., Bolinger, L., Scholz, T.D., Kayser, H., van der Geest, R., Reiber, J.H.C.: Anatomical modeling with fuzzy implicit surface templates: Application to automated localization of the heart and lungs in thoracic MR volumes. *Computer Vision and Image Understanding* 18(3), 218–230 (1999)
9. Lu, X., Georgescu, B., Jolly, M., Guehring, J., Young, A., Cowan, B., Littmann, A., Comaniciu, D.: Cardiac anchoring in mri through context modeling. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6361, pp. 383–390. Springer, Heidelberg (2010)
10. Tu, Z.: Probabilistic boosting-tree: Learning discriminative models for classification, recognition, and clustering. In: *Proc. ICCV*, pp. 1589–1596 (2005)
11. Viola, P., Jones, M.J.: Robust real-time face detection. *International Journal of Computer Vision* 57(2), 137–154 (2004)
12. Zheng, Y., Barbu, A., Georgescu, B., Scheuering, M., Comaniciu, D.: Fast automatic heart chamber segmentation from 3D CT data using marginal space learning and steerable features. In: *Proc. ICCV* (2007)

Multi-part Left Atrium Modeling and Segmentation in C-Arm CT Volumes for Atrial Fibrillation Ablation

Yefeng Zheng¹, Tianzhou Wang¹, Matthias John²,
S. Kevin Zhou¹, Jan Boese², and Dorin Comaniciu¹

¹ Image Analytics & Informatics, Siemens Corporate Research, Princeton, NJ, USA

² Healthcare Sector, Siemens AG, Forchheim, Germany
yefeng.zheng@siemens.com

Abstract. As a minimally invasive surgery to treat left atrial (LA) fibrillation, catheter based ablation uses high radio-frequency energy to eliminate potential sources of the abnormal electrical events, especially around the ostia of pulmonary veins (PV). Due to large structural variations of the PV drainage pattern, a personalized LA model is helpful to translate a generic ablation strategy to a specific patient's anatomy. Overlaying the LA model onto 2D fluoroscopic images provides valuable visual guidance during surgery. A holistic shape model is not accurate enough to represent the whole shape population of the LA. In this paper, we propose a part based LA model (including the chamber, appendage, and four major PVs) and each part is a much simpler anatomical structure compared to the holistic one. Our approach works on un-gated C-arm CT, where thin boundaries between the LA blood pool and surrounding tissues are often blurred due to the cardiac motion artifacts (which presents a big challenge compared to the highly contrasted gated CT/MRI). To avoid segmentation leakage, the shape prior is exploited in a model based approach to segment the LA parts. However, independent detection of each part is not optimal and its robustness needs further improvement (especially for the appendage and PVs). We propose to enforce a statistical shape constraint during the estimation of pose parameters (position, orientation, and size) of different parts. Our approach is computationally efficient, taking about 1.5 s to process a volume with $256 \times 256 \times 250$ voxels. Experiments on 469 C-arm CT datasets demonstrate its robustness.

1 Introduction

About 15% of all strokes (which is the third leading cause of death in the USA) are caused by atrial fibrillation (AF) [1]. As a widely used minimally invasive surgery to treat AF, the catheter based ablation procedure uses high radio-frequency energy to eliminate the sources of ectopic foci, especially around the ostia of the appendage and the pulmonary veins (PV). Automatic segmentation of the left atrium (LA) is important for the pre-operative assessment to identify

the potential sources of the abnormal electrical events [2,3]. However, there are large variations in the PV drainage patterns [4] and the most common variations (about 20-30% of the whole population) are extra right PVs and left common PV (where the two left PVs merge into one before joining the chamber). A personalized LA model can help to translate a generic ablation strategy to the specific patient's anatomy, thus making the ablation strategy more effective for this patient. It can also be overlaid onto 2D real-time fluoroscopic images to provide visual guidance during surgery (Fig. 1).

The previous LA segmentation methods can be roughly categorized as non-model based or model based approaches. The non-model based approaches [2,3] do not assume any prior knowledge of the LA shape and the whole segmentation procedure is purely data driven. For example, John and Rahn [2] proposed to use region growing to extract the highly contrasted blood pool in computed tomography (CT) or magnetic resonance imaging (MRI) data, followed by a split-and-merge procedure to identify the LA. The advantage of these methods is that they can handle structural variations of the PVs. However, such methods cannot provide the underlying anatomical information (e.g., which part of the segmentation is the left inferior PV). In practice, non-model based approaches work well on CT and MRI datasets, but they are not robust on challenging C-arm CT. The model based approaches exploit a prior shape of the LA (either in the form of an atlas [5,6] or a mean shape mesh [7]) to guide the segmentation. For example, Manzke *et al.* [7] built a mean shape of the combined structure of the LA chamber and PVs from a training set. With a prior shape constraint, they could avoid the leakage around weak or missing boundaries, which plagues the non-model based approaches. However, using one mean shape, it is difficult to handle the structural variations, e.g., the left common PV. Furthermore, they did not segment the LA appendage. The PV variations were addressed in [6] using multiple atlases, at the cost of extra computation time.

In this paper, we propose a fully automatic LA segmentation system on C-arm CT. Compared to conventional CT or MRI, the advantage of C-arm CT is that overlay of the 3D patient-specific LA model onto a 2D fluoroscopic image is straightforward and accurate since both the 3D and 2D images are captured on the same device within a short time interval. Normally, a non-electrocardiography-gated acquisition is performed to reconstruct a C-arm CT volume, therefore, it contains severe cardiac motion artifacts. For a C-arm with a small X-ray detector panel, part of the body may be missing in some 2D X-ray projections due to the limited field-of-view, resulting in significant artifacts around the margin of a reconstructed volume. In addition, there may be severe streak artifacts caused by various catheters inserted in the heart. These challenges are addressed using a model based approach, which also takes advantage of a machine learning based object pose detector and boundary detector [8].

Instead of using one mean model [7], the PV structural variations are addressed using a part based model, where the whole LA is split into the chamber, appendage, and four major PVs. Each part is a much simpler anatomical structure compared to the holistic one, therefore can be detected and segmented using

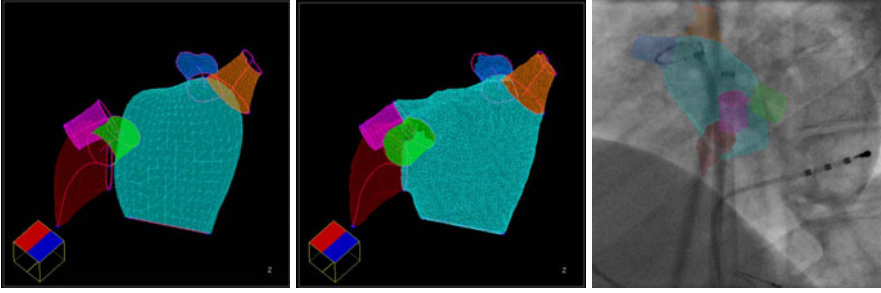


Fig. 1. Part based left atrium (LA) mesh model. **Left:** Meshes for the separate LA parts. **Middle:** Final consolidated mesh model. **Right:** Overlay of the model onto fluoroscopic images to provide visual guidance during surgery. Note: Cyan for the LA chamber, dark red for the appendage, green for the left inferior pulmonary vein (PV), magenta for the left superior PV, orange for the right inferior PV, and blue for the right superior PV.

a model based approach. In order to increase robustness, we detect the most reliable structure (the LA chamber) and use it to constrain the detection of other parts (the appendage and PVs). Experiments show that it is better to treat the LA chamber and appendage as a single object to improve the detection robustness of the appendage. Due to the large variations, the relative position of the PVs to the LA chamber varies quite a lot. A statistical shape model is used to enforce a proper constraint during the estimation of the PV pose parameters (position, orientation, and size). The proposed method is computationally efficient, taking about 1.5 s to process a volume with $256 \times 256 \times 250$ voxels. Extensive experiments on 469 un-gated C-arm CT datasets demonstrate the robustness of the proposed method.

2 Multi-part Left Atrium Modeling and Segmentation

2.1 Part Based Left Atrium Model

As shown in the left column of Fig. 1, our part based LA model includes the LA chamber body, appendage, and four major PVs. The shape of the appendage is close to a tilted cone and the PVs have a tubular structure. For AF ablation, physicians only care about a proximal PV trunk, therefore, we only detect a trunk of 20 mm in length, originating from its ostium. Each LA part is a much simpler anatomical structure compared to the holistic one, therefore can be detected and segmented using a model based approach. We use marginal space learning (MSL) [8] to detect and segment each part. MSL is an efficient method to estimate the position, orientation, and size of the object in a 3D volume. After automatic object pose estimation, a mean shape is aligned with the pose as an initial mesh. A machine learning based boundary detector [8] is used to guide the boundary evolution. For each LA part, we have an MSL based pose detector and a learning based boundary detector. Interested readers are referred to [8] for

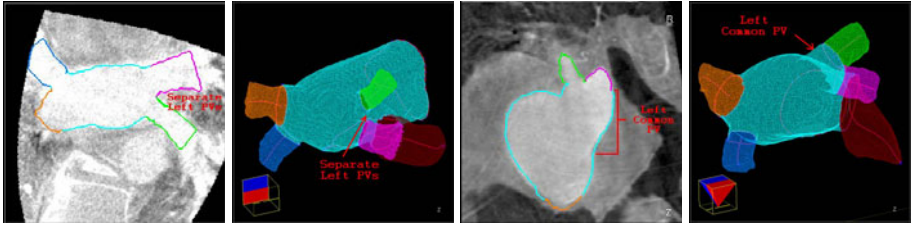


Fig. 2. Pulmonary vein (PV) segmentation results on two datasets. **Left:** A patient with separate left inferior (green) and superior (magenta) PVs. **Right:** A patient with the left common PV.

more details of MSL. After detecting/segmenting all parts, we merge them into a consolidated mesh (the middle column of Fig. 1), using the technique presented in Section 2.3.

2.2 Constrained Detection of LA Parts

Compared to the holistic approach [7], the part based approach can handle large structural variations. The MSL based detection/segmentation works well for the LA chamber. However, independent detection of other parts is not robust, either due to the low contrast (appendage) or the small object size (PVs). In C-arm CT, the appendage is particularly difficult to detect since the appendage is a pouch without outlet and the blood flow is slow inside the appendage, preventing the complete filling of contrast agent. In many datasets, the appendage is only barely visible. The MSL detector may pick the neighboring left superior PV, which often touches the appendage and has higher contrast. However, the relative position of the appendage to the chamber is quite consistent. The best performance is achieved by treating the appendage and chamber as a consolidated object. One MSL based pose detector is trained to detect the combined object.

Through comparison experiments, we found neither a holistic approach nor independent detection worked for the PVs. In this paper, we propose to enforce a statistical shape constraint [9] in PV detection. The point distribution model (PDM) is often used to enforce the statistical shape constraint among a set of landmarks. The shape variation is decomposed into orthogonal deformation modes through principal component analysis (PCA). A deformed shape is projected into a low dimensional deformation subspace to enforce a statistical shape constraint. For each PV, an MSL detector can estimate nine pose parameters, i.e., three position parameter (T_x, T_y, T_z) , three orientation Euler angles (O_x, O_y, O_z) , and three anisotropic scaling parameters (S_x, S_y, S_z) . Different to the conventional PDM, we also want to enforce the constraint among the estimated orientation and size of PVs. One solution is to stack all PV pose parameters into a big vector to perform PCA. However, the position and orientation parameters are measured in different units. If not weighted properly, the extracted deformation modes may be dominated by one category of transformation. Furthermore, the Euler angles are periodic (with a period of 2π), which

prevents the application of PCA. Boisvert *et al.* [10] proposed to build a shape model on a Riemannian manifold on which we have an intrinsic measurement of the orientation distance. However, they still need to heuristically assign a proper weight to the distance in translation w.r.t. the distance in orientation.

In this work, we use a new presentation of the pose parameters to avoid the above problems. The object pose can be fully represented by the object center T together with three scaled orthogonal axes. Alternative to the Euler angles, the object orientation can be represented as a rotation matrix $R = (R_x, R_y, R_z)$ and each column of R defines an axis. The object pose parameters can be fully represented by a four-point set (T, V_x, V_y, V_z) , where

$$V_x = T + S_x R_x, \quad V_y = T + S_y R_y, \quad V_z = T + S_z R_z. \quad (1)$$

The pose of each PV is represented as four points. Besides the constraint among the PVs, we also add the already detected LA chamber center and appendage center to stabilize the detection. In total, we get a set of 18 points.

In our experiments, after detecting the position, orientation, size of the PVs, we project their poses into a subspace with eight dimensions (which explains about 75% of the total variation) to enforce a statistical shape constraint. After enforcing a statistical shape constraint, the PV center is given by point \hat{T} . We can recover the orientation (\hat{R}) and scale (\hat{S}) by simple inversion of Eq. (1). However, the estimate \hat{R} is generally not a true rotation matrix ($\hat{R}^T \hat{R} \neq I$). We want to find the nearest rotation matrix R_o to minimize the sum of squares of elements in the difference matrix $R_o - \hat{R}$, which is equivalent to

$$R_o = \arg \min_R \text{Trace}((R - \hat{R})^T (R - \hat{R})), \quad (2)$$

subject to $R_o^T R_o = I$. Here, $\text{Trace}(\cdot)$ is sum of the diagonal elements. The optimal solution [11] is given by

$$R_o = \hat{R}(\hat{R}^T \hat{R})^{-1/2}. \quad (3)$$

2.3 Mesh Consolidation

After constrained detection and segmentation, we get six separate meshes (the LA chamber, appendage, and four PVs), as shown in Fig. 1. There may be gaps and/or intersections among different meshes. The physicians prefer a consolidated mesh with different anatomical structures labeled with different colors. We first project the proximal rim of a PV or appendage along the centerline onto the LA chamber to eliminate the gaps among different mesh parts (Fig. 3b). Now, the part meshes are fully connected. However, the mesh intersections may still be present since a piece of the PV mesh may lie inside the chamber. It is complicated to work directly on the meshes. Instead, we convert the meshes to a volume mask (Fig. 3c), and generate a new mesh (Fig. 3d) from the volume mask using the marching cubes algorithm [12]. Since PV/appendage meshes are connected to the LA chamber in a pure geometric operation, the region around the ostia is not segmented accurately. After converting the meshes to a volume mask, we perform constrained region growing a few layers to improve the segmentation accuracy around the ostia of appendage and PVs.

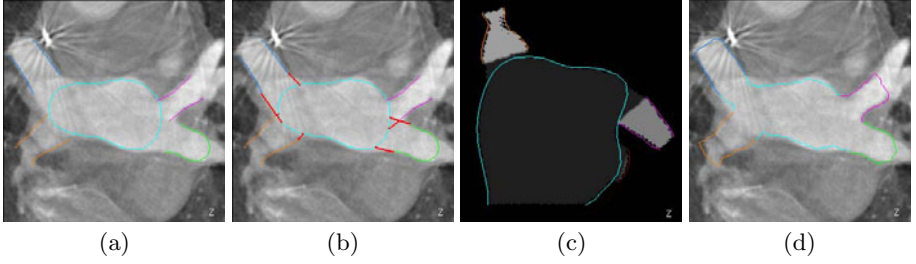


Fig. 3. Generating a consolidated left atrium (LA) mesh from separate parts. (a) Separate meshes of the pulmonary veins (PV) and LA chamber. (b) After connecting the PV meshes to the LA chamber. The added mesh pieces are shown in red. (c) Volume mask generated from the meshes. (d) Consolidated mesh.

Table 1. Left atrium segmentation errors (based on four-fold cross validation) on C-arm CT datasets with 253 large volumes and 216 small volumes. The symmetric surface-to-surface errors, measured in millimeters (mm), are reported.

	Large Volumes						Small Volumes					
	Holistic		Independent		Proposed		Holistic		Independent		Proposed	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median
LA Chamber	1.82	1.69	1.76	1.60	1.69	1.49	2.01	1.77	1.87	1.70	1.81	1.62
Appendage	3.28	2.65	3.49	2.49	2.89	2.18	3.41	2.67	3.77	3.01	2.93	2.49
Left Inf. PV	3.08	2.49	2.01	1.59	1.88	1.54	4.12	2.89	2.35	1.66	2.01	1.76
Left Sup. PV	2.77	2.21	1.82	1.42	1.83	1.39	3.48	2.44	2.56	1.54	1.88	1.49
Right Inf. PV	3.77	2.81	2.15	1.44	1.89	1.43	3.93	3.09	2.34	1.66	2.10	1.74
Right Sup. PV	2.68	2.29	1.74	1.35	1.61	1.40	3.08	2.50	1.72	1.35	1.62	1.31
Whole Mesh	2.00	1.86	1.71	1.58	1.63	1.51	2.17	1.98	1.79	1.63	1.71	1.55

3 Experiments

We collected 469 C-arm CT datasets, scanned by Siemens Axiom Artis zee C-arm systems at 17 clinical sites from Europe and the USA. Among them, 253 datasets were scanned with large X-ray detector panels ($30 \times 40 \text{ cm}^2$) with an isotropic volume resolution of 0.30 mm^3 . The other 216 datasets were scanned with small X-ray detectors ($20 \times 20 \text{ cm}^2$) with a volume resolution of 0.18 mm^3 . Due to the limited field-of-view of a small X-ray detector, the reconstructed volumes may contain significant artifacts, especially around the volume margin. Due to the heterogeneity of the datasets, we train two separate systems for the large and small volumes, respectively.

A four-fold cross validation is performed to evaluate our algorithm and the segmentation accuracy is measured using the symmetric surface-to-surface distance. Table 1 shows segmentation errors of the consolidated mesh using various approaches. For a mesh part, we measure the distance to the corresponding part in the ground truth, therefore both the segmentation and mesh part labeling errors are penalized. At the bottom row of Table 1, we also list the whole mesh segmentation error without considering the mesh part labeling. If we treat the whole LA mesh as a holistic object, the segmentation errors are large, especially

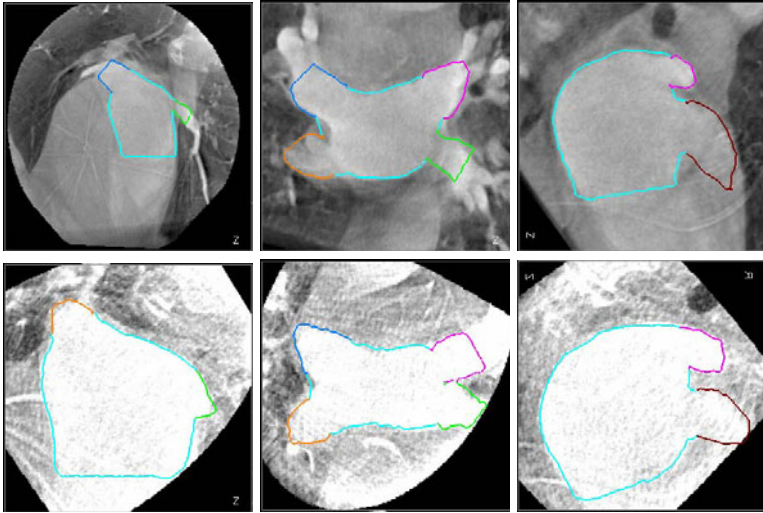


Fig. 4. The left atrium segmentation results on a large volume (top row) and a small volume (bottom row) with different colors for different mesh parts. Three different views are shown for each volume.

for the PVs due to the structural variations. Independent detection of the parts can significantly improve the PV segmentation accuracy. However, the LA appendage segmentation is deteriorated. Using the proposed method, we improve the segmentation for all LA parts. Compared to the independent detection, on average, we achieve about 10% reduction in mean error for PVs (up to 26.6% reduction for the left superior PV on small volumes). Inferior PVs have larger errors than superior PVs since they are more likely to be cut by the volume borders.

We cannot directly compare our segmentation accuracy with those reported in the literature due to the difference in imaging modalities, datasets, and LA models. There is no quantitative evaluation available in [2,3]. To the best of our knowledge, there is only one previous work [7] on the LA segmentation on C-arm CT, which was validated on large volumes only and did not segment the more challenging LA appendage. After excluding the LA appendage, we achieve an overall segmentation error of 1.50 mm for large volumes and 1.63 mm for small volumes. In [7], a mean segmentation error of 1.3 mm is reported on 33 datasets (which is significantly smaller than ours). Their segmentation is initialized with a holistic mean shape. It is not clear if patients with the left common PV are included in their limited datasets, since a holistic approach has difficulty to handle such structural variations. We also notice that the mesh triangles near the distal bounds of the truncated PVs (9.4% of all triangles) are excluded from the error measurement, however, we include all mesh triangles in evaluation.

The proposed method is computationally efficient, taking about 1.5 s (on a computer with quad-core 2.33 GHz CPUs and 3 GB memory) to process a

volume with $256 \times 256 \times 250$ voxels. It compares favorably with the previous methods, e.g., 5 s in [5], 5-45 s in [2], 30 s in [7], and more than two hours in [6].

4 Conclusion

We proposed a part based model for the automatic segmentation of the LA (including chamber, appendage, and four major PVs) in C-arm CT volumes, which can handle structural variations elegantly. Extensive experiments on 469 ungated C-arm CT datasets demonstrated the robustness and efficiency of the proposed method. As an extension, we are currently working on the automatic segmentation of the extra right PVs if they are present. Our system is re-trainable, therefore can be extended to different medical imaging modalities, e.g., MRI and conventional CT.

References

1. Lloyd-Jones, D., Adams, R., Carnethon, M., et al.: Heart disease and stroke statistics – 2009 update. *Circulation* 119(3), 21–181 (2009)
2. John, M., Rahn, N.: Automatic left atrium segmentation by cutting the blood pool at narrowings. In: Duncan, J.S., Gerig, G. (eds.) *MICCAI 2005*. LNCS, vol. 3750, pp. 798–805. Springer, Heidelberg (2005)
3. Karim, R., Mohiaddin, R., Rueckert, D.: Left atrium segmentation for atrial fibrillation ablation. In: *Proc. of SPIE Medical Imaging* (2008)
4. Marom, E.M., Herndon, J.E., Kim, Y.K., McAdams, H.P.: Variations in pulmonary venous drainage to the left atrium: Implications for radiofrequency ablation. *Radiology* 230, 824–829 (2004)
5. Karim, R., Juli, C., Lawes, L.M., Kanangaratnam, P., Davies, D.W., Peters, N.S., Rueckert, D.: Automatic segmentation of left atrial geometry from contrast-enhanced magnetic resonance images using a probabilistic atlas. In: Camara, O., Pop, M., Rhode, K., Sermesant, M., Smith, N., Young, A. (eds.) *STACOM 2010*. LNCS, vol. 6364, pp. 134–143. Springer, Heidelberg (2010)
6. Depa, M., Sabuncu, M.R., Holmvang, G., Nezafat, R., Schmidt, E.J., Golland, P.: Robust atlas-based segmentation of highly variable anatomy: Left atrium segmentation. In: Camara, O., Pop, M., Rhode, K., Sermesant, M., Smith, N., Young, A. (eds.) *STACOM 2010*. LNCS, vol. 6364, pp. 85–94. Springer, Heidelberg (2010)
7. Manzke, R., Meyer, C., Ecabert, O., Peters, J., Noordhoek, N.J., Thiagalingam, A., Reddy, V.Y., Chan, R.C., Weese, J.: Automatic segmentation of rotational X-ray images for anatomic intra-procedural surface generation in atrial fibrillation ablation procedures. *IEEE Trans. Medical Imaging* 29(2), 260–272 (2010)
8. Zheng, Y., Barbu, A., Georgescu, B., Scheuring, M., Comaniciu, D.: Four-chamber heart modeling and automatic segmentation for 3D cardiac CT volumes using marginal space learning and steerable features. *IEEE Trans. Medical Imaging* 27(11), 1668–1681 (2008)
9. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active shape models—their training and application. *Computer Vision and Image Understanding* 61(1), 38–59 (1995)

10. Boisvert, J., Pennec, X., Labelle, H., Cheriet, F., Ayache, N.: Principal spine shape deformation modes using Riemannian geometry and articulated models. In: Perales, F.J., Fisher, R.B. (eds.) AMDO 2006. LNCS, vol. 4069, pp. 346–355. Springer, Heidelberg (2006)
11. Horn, B.K.P.: Closed form solution of absolute orientation using unit quaternions. *Journal of the Optical Society A* 4(4), 629–642 (1987)
12. Lorensen, W.E., Cline, H.E.: Marching cubes: A high resolution 3D surface construction algorithm. *Computer Graphics* 21(4), 163–169 (1987)

Accurate and Robust Fully-Automatic QCA: Method and Numerical Validation

Antonio Hernández-Vela^{1,2}, Carlo Gatta^{1,2}, Sergio Escalera^{1,2},
Laura Igual^{1,2}, Victoria Martín-Yuste³, and Petia Radeva^{1,2}

¹ Dept. MAIA, Universitat de Barcelona, Gran Via 585, 08007 Barcelona, Spain
² Centre de Visió per Computador, Edifici O, Campus UAB, 08193 Bellaterra, Spain
{ahernandez, cgatta, sescalera, ligual, petia}@cvc.uab.cat
³ Institut Clinic del Torax, Hospital Clinic Barcelona,
Villarroel 170, 08036 Barcelona, Spain
27700vmy@comb.cat

Abstract. The Quantitative Coronary Angiography (QCA) is a methodology used to evaluate the arterial diseases and, in particular, the degree of stenosis. In this paper we propose AQCA, a fully automatic method for vessel segmentation based on graph cut theory. Vesselness, geodesic paths and a new multi-scale edgeness map are used to compute a globally optimal artery segmentation. We evaluate the method performance in a rigorous numerical way on two datasets. The method can detect an artery with precision $92.9 \pm 5\%$ and sensitivity $94.2 \pm 6\%$. The average absolute distance error between detected and ground truth centerline is 1.13 ± 0.11 pixels (about 0.27 ± 0.025 mm) and the absolute relative error in the vessel caliber estimation is 2.93% with almost no bias. Moreover, the method can discriminate between arteries and catheter with an accuracy of 96.4%.

Keywords: Vessel segmentation, centerline extraction, QCA, GraphCut.

1 Introduction

The enhancement and segmentation of tubular structures and/or vessel-like structures is a prolific topic in the medical imaging research; many methods, exploiting photometric and structural properties of tubular structures have been proposed (see *e.g.* a complete review of recent methodologies for CTA segmentation in [11]). Nonetheless, in the case of vessel segmentation in angiography sequences, the problem is still very hard; highly reliable, fully automatic methods are not established yet [5]. Finally, accurate segmentation is still a hot topic and far from being solved, as demonstrated by the excellent scale selection method proposed in [10]. An extensive overview of different methods for vessel extraction is provided in [7]. Recently, an interesting approach to vessel segmentation has been proposed in [12], which fuses local features with local directional information; unfortunately, authors do not provide a quantitative evaluation of their method. Nonetheless, most works are based on local image analysis to extract vessels or employ an a-priori model to help vessel extraction. In contrast, graph cut (GC) technique is an optimal segmentation tool that combines local and contextual image information analysis by modeling relations between neighboring pixels. The GC algorithm [28] has been used in many computer vision problems and, in particular, it can

be applied to binary-segmentation of images, obtaining a solution which corresponds to the global minimum of an energy function. The goodness of the solution depends on the suitability of the energy terms and their reliable computation.

In this paper, we use GC theory to obtain a globally optimal segmentation of the coronary tree in angiography images. Differently than other methods, we are interested in accurate detection of both the centerline and the vessel borders. To this aim, we propose a novel energy function tailored to the artery segmentation problem. The energy takes into account (1) the vessel local appearance, using any vesselness measure, (2) the local connectivity to other vessel regions, using geodesic paths and, (3) a measure of edginess based on a novel multi-scale version of the adaptive Canny detector [4]. Moreover, we propose a machine learning based approach to automatically detect the catheter; providing a methodology that is far more general than the one in [12]. Finally, we propose two datasets, which allow the quantitative evaluation of the method in terms of ability to detect the artery (Precision and Sensitivity), errors in the centerline detection and caliber estimation, and ability to discriminate between arteries and catheters¹. To the best of our knowledge, there is no such dataset available in the literature. In next section, we provide a brief introduction to GC, then detail our contributions.

2 Method

2.1 Graph Cut

Let us define $\mathcal{X} = (\mathbf{x}_1, \dots, \mathbf{x}_i, \dots, \mathbf{x}_{|\mathcal{P}|})$ the set of pixels for a given angiography gray-scale image I ; $\mathcal{P} = (1, \dots, i, \dots, |\mathcal{P}|)$ the set of indexes of I ; \mathcal{N} the set of unordered pairs $\{i, j\}$ of neighboring pixels of \mathcal{P} under a 4-(8-) neighborhood system, and $L = (L_1, \dots, L_i, \dots, L_{|\mathcal{P}|})$ a binary vector whose components L_i specify assignments to pixels $i \in \mathcal{P}$. Each L_i can be either “vess” or “back” indicating if it belongs to vessel or background, respectively. GC formulation [2] defines the cost function $E(L)$, which describes soft constraints imposed on boundary and region properties of L as $E(L) = U(L) + \lambda B(L)$. The unary term is denoted with $U(L) = \sum_{i \in \mathcal{P}} U_i(L_i)$, the boundary term with $B(L) = \sum_{\{i,j\} \in \mathcal{N}} B_{\{i,j\}} \Omega(L_i = L_j)$, where the characteristic function $\Omega(L_i = L_j)$ is 0 if $L_i \neq L_j$ and 1, otherwise. The unary term $U(L)$ is defined assuming that individual penalties for assigning pixel i to “vess” and “back”, (i.e. U_i (“vess”) and U_i (“back”)) are given by vessel and background models. The term $B(L)$ comprises the boundary properties of segmentation L . Any $B_{\{i,j\}} \geq 0$ should be interpreted as a penalty for a discontinuity between i and j . Finally, the coefficient $\lambda \in \mathbb{R}^+$, $\lambda \geq 0$ specifies the relative importance of the boundary term against the unary term. GC algorithm imposes hard constraints on the segmentation result by means of the definition of seed points where labels are predefined and can not be modified. The subsets $\mathcal{V} \subset \mathcal{P}$, $\mathcal{B} \subset \mathcal{P}$, $\mathcal{V} \cap \mathcal{B} = \emptyset$ denote the subsets of vessel and background seeds, respectively.

Boykov et al. [2] show how to efficiently compute the global minimum of $E(L)$ among all segmentations L satisfying the hard constraints $\forall i \in \mathcal{V}, L_i = \text{“vess”}$, $\forall i \in \mathcal{B}, L_i = \text{“back”}$, using a minimum cut algorithm on a certain graph defined by nodes

¹ The datasets and the evaluation methodology will be provided under request.

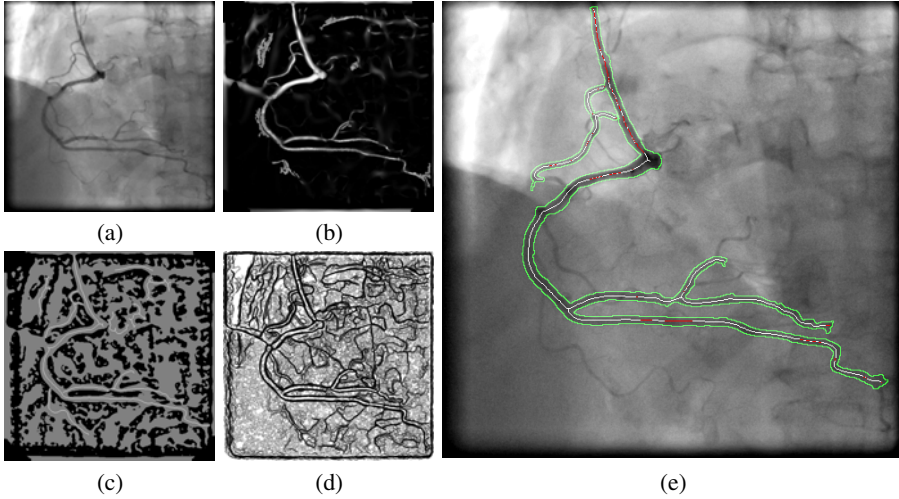


Fig. 1. AQCA approach: (a) Input image, (b) Unary potential: Vesselness, (c) Vessel-Background seeds (in white and black), (d) Boundary potential: Multi-scale edgeness map, (e) Final segmentation, centerline estimation, and catheter detection (in green, white, and red).

and edges being image pixels and pixel relations, respectively. We use the algorithm presented in [3] for computing the minimum cut.

2.2 Vessel Segmentation Algorithm

In this section, we describe in detail the seed initialization and the definition of unary and boundary potentials.

Seed initialization. In order to achieve a fully automatic methodology, we exploit the inherent structure of vessels to define vessel seeds based on valleys, and background seeds based on low probabilities of the vesselness image [6]. In particular, vessel seeds \mathcal{V} correspond to those pixels corresponding to the highest responses on a multilocal valley detector with structure tensor, namely \mathcal{S}_t , as described in [9], $\mathcal{V} = \{i | \mathcal{S}_{t,i} > \Theta_v\}$, where $\mathcal{S}_{t,i}$ is the valley response at pixel i , and Θ_v is a sensitivity valley threshold. The background seeds \mathcal{B} are the pixels corresponding to low probabilities in the vesselness image V , $\mathcal{B} = \{i | V_i \leq \Theta_b\}$, where Θ_b is a sensitivity vesselness threshold and the vesselness measure at pixel i , V_i , is computed as in [6]. Fig. 1(c) shows the selection of \mathcal{V} and \mathcal{B} seeds for the input image in Fig. 1(a).

Unary term. We define the vessel and background models using the vesselness map V . However, some vessel regions (especially those corresponding to bifurcations), can contain low vessel probability. To avoid this problem, we introduce the computation of geodesic paths among vessels seeds. We initialize the unary potentials at each pixel i as $U_i(\text{“vess”}) = -\ln(p(L_i = \text{“vess”}))$, $U_i(\text{“back”}) = -\ln(p(L_i = \text{“back”}))$. The probability of a pixel to be marked as “vess” is computed using the vesselness-geodesic measure VG , $p(L_i = \text{“vess”}) = VG(\mathbf{x}_i)$ and the opposite probability as

$p(L_i = \text{“back”}) = 1 - p(L_i = \text{“vess”})$. In particular, the map VG is computed for pixel i as the maximum between the vesselness value and the inverse of the geodesic distance,

$$VG_i = \max \left(V_{o,i}, \frac{1}{\max \left(\frac{D_i + \mu(D)}{1}, \frac{1}{D_i + \mu(D)} \right)} \right),$$

where D and $\mu(D)$ correspond to the geodesic distance map and its mean, respectively. D is computed as explained in the following subsection.

Geodesic map. Given an arbitrary parameterized discrete path $\Gamma = \{i, \dots, j\}$ defined by $|\Gamma| = R$ pixels, we define the geodesic distance D of Γ as,

$$D(\Gamma) = m(\|\nabla I(\Gamma)\|^2) \left(\sum_{i=1}^{R-1} \frac{\|\nabla I_i\|^2}{R} \right), \tag{1}$$

where the quantity $\|\nabla I_i\|$ is a finite difference approximation of the image gradient between points $(\mathbf{x}_i, \mathbf{x}_{i+1})$, and the function $m(\mathbf{z})$ represents the maximum variance of the R -dimensional vector \mathbf{z} , $m(\mathbf{z}) = \max_{i,j} |\mathbf{z}_i - \mathbf{z}_j|, i, j \in \Gamma$. The measure defined in Eq.(1) is normalized by the length of the path, allowing any path length to be considered. However, it is penalized by the maximum variance of image gradients within the path. The selected geodesic path is given by, $\Gamma_{\{i,j\}}^* = \operatorname{argmin}_{\Gamma \in \mathcal{C}_{\{i,j\}}} D(\Gamma)$, and its distance measure is $d(i, j) = \min_{\Gamma \in \mathcal{C}_{\{i,j\}}} D(\Gamma)$, being $\mathcal{C}_{\{i,j\}}$ the set of all possible paths between points i and j using short-path algorithm. We proceed as follows: after computing the partial path $\Gamma_{\{i,j\}}^*$, in order to select the next path point $j^* \in \mathcal{G}_j$, where \mathcal{G}_j is the set of 8-neighbor of j , we use the following criterion, $j^* = \operatorname{argmin}_{\ell \in \mathcal{G}_j} (d(i, j) + \sigma_{\{j,\ell\}} D(\Gamma_{\{j,\ell\}}))$, where $\sigma_{\{j,\ell\}}$ is the variance between j and ℓ . Once the next point has been selected, we continue the path only if $d(i, j^*) < \Theta_d$, where $\Theta_d = 0.05$ is an XRay-dependent empirically set threshold. Since different geodesic maps can be found for different initialization pixels j , the geodesic map, for each pixel i , is computed as $D_i = \min_{j \in \Gamma} D(\Gamma_{\{i,j\}})$. These pixels are the centroids of a k -means clustering over the vessel seeds. An example of the VG map is shown in Fig. 1(b).

Boundary term. We propose an image-dependent multi-scale edgeness measure. First, we run the canny edge detector algorithm on the observed image at different threshold levels. Then, we compute the edge probability at each pixel by the linear average of the edge thresholds and for different scales as follows, $J_i^* = \min_j \frac{1}{n} \sum_{k=1}^n J_{p,\gamma_k,\sigma_j}$, where J_{p,γ_k,σ_j} is the binary edge map using the threshold $\gamma_k \in [0.02, 0.03, \dots, 0.3]$ and scale $\sigma_j \in [0.5, 1, \dots, 5]$ for pixel i . If pixel i is labeled as an edge pixel for most of the threshold levels at a significant scale, it has a high probability of being an edge pixel. The final boundary potential over the multi-scale edgeness map is computed as $B_{\{i,j\}} = J_i^*$. An example is shown in Fig. 1(d).

2.3 Postprocessing

We perform a post-filtering step consisting in keeping only the biggest connected component in the final segmentation. The main aim of this step is to get rid of possible false

positive (FP) regions that could be introduced by the seed initialization. An example of the final segmentation is shown in green in Fig. 1(e).

The centerline (CR) is extracted as follows: given the binary segmentation $L(\mathbf{x}_i)$, we compute its distance map $M(\mathbf{x}_i)$. Then, a non-maxima suppression is applied to find local maxima and a classic ridge transversal method is applied to connect the local maxima. The ridge transversal stops when it finds another centerline or it exits the segmented area. Fig. 1(e) shows an example of an extracted centerline in white.

Vessel caliber is estimated by applying a local Laplacian of Gaussian (LoG) filtering at CRs locations at different scales. The scale space computed using $\sigma^2 \text{LoG}(x, y; \sigma)$ has a minimum at $\sigma = w/2$, where w is the width of the ridge.

Catheter detection. By merely its appearance, the catheter is not easily distinguishable from arteries. This causes that the proposed segmentation method tends to segment the catheter as an artery. In order to detect the catheter, from each point of the centerline path we extract: (1) its position \mathbf{x} , (2) its curvature $K(\mathbf{x})$, (3) its angular direction $\alpha(\mathbf{x})$, and (4) its caliber $C(\mathbf{x})$. A classification is performed in a point wise way using a Bayesian classifier. Being $c = \{\text{“cat”}, \text{“vess”}\}$ the catheter or artery class, we modeled (1) $p(\mathbf{x}|c)$ using a Kernel Density Estimator with $\sigma_{\text{KDE}} = 15\text{pixels}$, (2) $p(\log(\epsilon + K(\mathbf{x}))|c)$ using a Gaussian Mixture Model² (10 Gaussians), (3-4) $p(\alpha(\mathbf{x})|c)$ and $p(C(\mathbf{x}|c))$ as two discrete histograms. In Fig. 1(e) an example of catheter detection is shown in red.

The time complexity of the algorithm most expensive part is $O(|\mathcal{P}| + \frac{|\mathcal{N}|}{|\mathcal{P}|} \log |\mathcal{P}|)$.

3 Validation

Material. We defined two datasets: DS1 and DS2. DS1 is formed by 20 images acquired with a single plane Philips INTEGRIS Allura Flat Detector, of RCAs, where three experts have blindly annotated the centerlines. The experts had to annotate the centerlines with different labels: “vess”: the arteries that potentially can present a clinical interest (with a caliber of, at least, 1mm); “don’t care”: all other arteries in the image, and “cat”: the catheter guide. DS2 is formed by 31 images from 27 patients, acquired with a SIEMENS Artis zee, of 10 RCAs, 10 LADs, and 11 Cxs. Two experts blindly segmented a total of 41 lesions (12 LADs, 13 Cxs and 16 RCAs) assisted by a semi-automatic method (QCA-CMS Version 6.0, MEVIS). The experts were asked to manually correct unsatisfactory segmentations.

Methods. We compare our proposed method (AQCA) against a classic ridge transversal centerline extraction method (RT), and the state of the art GC method. Furthermore, we also compare the obtained results with the inter-observer (IO) variability of the experts ground truth (GT). In the case of AQCA and GC, parameters Θ_v , Θ_b , and λ are tuned via cross-validation over DS1. Given N patients, the tuning is performed using a Leave One Patient Out (LOPO) methodology by maximizing $0.5P + 0.5S$ in order to provide a balance between precision (P) and sensitivity (S).

² The curvature distribution was far from being Gaussian, so we applied the logarithm to “Gaussianize” it.

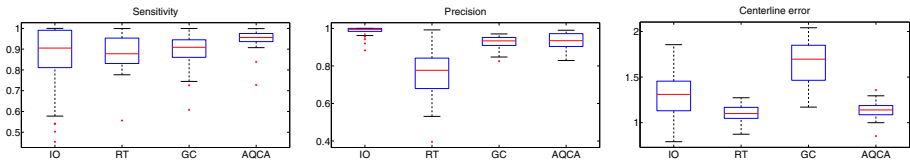


Fig. 2. Centerline evaluation results for DS1

Validation protocol. For the evaluation of centerline detection, we compute P , S , localization error (\mathcal{E}_L) on DS1. Additionally we evaluate the caliber estimation on DS2, measuring the mean signed relative error and mean absolute relative error. Finally, the catheter detection is evaluated in terms of P , S , and accuracy.

Centerline evaluation: To evaluate the CR detection, we computed P , S , and \mathcal{E}_L considering Θ_c , which defines the maximal distance between the GT and the detected CR. To compute the S , we check for every CR point in the GT if there exists a detected CR point in a distance smaller than Θ_c ; if this happens, this point is considered as a True Positive (TP). Similarly, P is computed by checking the detected points instead than the ones in the GT. Θ_c has been set to 5 pixels to allow large \mathcal{E}_L errors.

Caliber evaluation: Vessel caliber evaluation was performed over DS2. We approximated two cubic splines to the borders annotated by the experts. Using these splines, we determined the CR and extracted the caliber for each point [10]. For each point in the GT CR, we localized the nearest point in the detected CR and evaluated the caliber estimation error using the Euclidean distance. We computed the signed error $\Delta D_c = D_c - D_c^*$, where D_c^* is the ground truth caliber, in millimeters. Finally, we defined the average absolute and signed relative errors $\frac{|\Delta D_c|}{D_c^*}$ and $\frac{\Delta D_c}{D_c^*}$, respectively.

4 Results

Figure 2 shows S , P , and \mathcal{E}_L for the IO variability, the RT, GC, and AQCA for the dataset DS1³. The RT method has the lowest S and a very low P , while the \mathcal{E}_L is very low; this confirms that the vesselness measure is well suited to accurately detect the CR, but it has the disadvantage to produce many FPs as confirmed by the low P . A basic GC approach increases both S and P while \mathcal{E}_L is increased due to inaccurate border detection using a gray-level based boundary term. Our proposed method shows the highest S and P , and a \mathcal{E}_L that is very close to RT, while actually detecting more vessel pixels than both RT and GC (higher S , less false negatives). It is also interesting to note that the proposed method has a lower P than the IO variability but higher S : this means that the proposed method still produces some FPs but tends to detect clinically relevant arteries in a way that is the “average” of the observers. Figure 3 shows scatter plots of the caliber estimation on dataset DS2 for the IO variability, the basic GC, and AQCA, respectively. The gray dashed curve shows the density of points w.r.t. the caliber. It can

³ The results for both datasets are available at www.cvc.uab.es/~ahernandez/AQCA.zip.

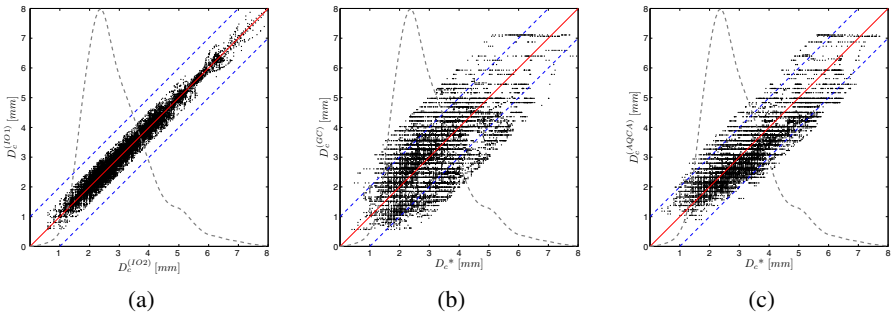


Fig. 3. GT and estimated calibers in scatter plots for (a) IO variability, (b) GC, and (c) AQCA

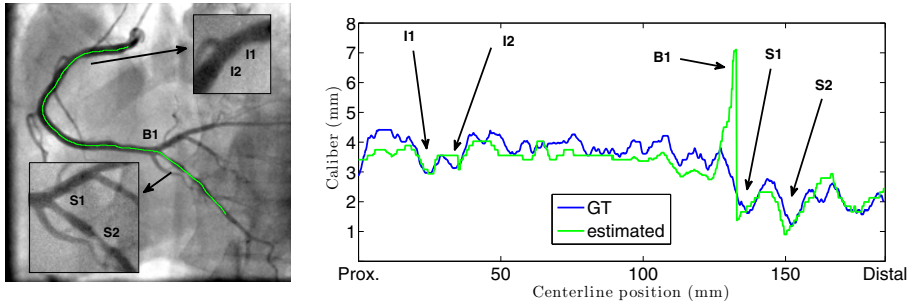


Fig. 4. Extracted centerline and GT and estimated caliber along the extracted centerline

be noticed that the two experts show a very high correlation. This is due to the fact that the segmentation has been performed thanks to a computer assisted method: if there is a bias in the measurement, it is equal for both experts. The basic GC method performs badly, as confirmed by the large average relative error of 6.04%, while the proposed method performs much better having an average relative error of 2.93%. It has to be noted that, while GC present a bias of -1.6% , our method presents almost no bias (-0.035%). Moreover, for the calibers which are more frequent in X-Ray angiography, our proposed method has a very high correlation with the ground truth data. Figure 4 shows an example of AQCA result: the estimated caliber from proximal to distal positions accurately follows the GT. At proximal position the method correctly detects two clinically irrelevant caliber variations (I1 and I2). Then, at the bifurcation B1, our method estimates the bifurcation “caliber” that is obviously not relevant, and can be easily detected as an outlier. Just after the bifurcation, the proposed method accurately measures the vessel caliber at the stenosis S1 and S2. Finally, we collected ground truth data (56406 samples) from DS1 and evaluated the catheter detection methodology on a leave-one-patient-out methodology, obtaining an average S of 70.9%, P of 90.1% and accuracy of 96.4%.

5 Conclusion

We presented AQCA, a novel segmentation method for X-Ray angiography images that takes into account vessel appearance, artery tree continuity, and borders appearance within the graph-cut theory. The algorithm has been tested on two new data sets.

Despite it has been tuned on DS1, it provided excellent results on DS2, showing the inherent robustness of the approach. While being applied on accurate QCA, the method could be profitably used as preprocessing for non-rigid multimodal registration algorithms. Moreover, it can be easily adapted to detect tubular structures in other kind of images. Future lines of research encompass the use of an high order potential to deal with irregularity at bifurcations and crossings; a supervised method to optimize the seed selection; an intelligent procedure to semantically tag the artery in order to obtain an automated QCA report for all relevant branches.

Acknowledgments. This work has been supported in part by the projects: La Marató de TV3 082131, TIN2009-14404-C02, and CONSOLIDER-INGENIO CSD 2007-00018. The work of C. Gatta is supported by a Beatriu de Pinos Fellowship.

References

1. Benmansour, F., Cohen, L.: Tubular structure segmentation based on minimal path method and anisotropic enhancement, vol. 92, pp. 192–210. Springer, Netherlands (2011)
2. Boykov, Y., Funka-Lea, G.: Graph cuts and efficient n-d image segmentation, vol. 70, pp. 109–131. Kluwer Academic Publishers, Hingham (2006)
3. Boykov, Y., Kolmogorov, V.: An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision, vol. 26, pp. 359–374. IEEE Computer Society, Los Alamitos (2001)
4. Candemir, S., Akgul, Y.: Adaptive regularization parameter for graph cut segmentation. In: Campilho, A., Kamel, M. (eds.) ICIAR 2010. LNCS, vol. 6111, pp. I:117–I:126. Springer, Heidelberg (2010)
5. Fallavollita, P., Cheriet, F.: Towards an automatic coronary artery segmentation algorithm. In: EMBS 2006, pp. 3037–3040. IEEE Computer Society, Los Alamitos (August 2006)
6. 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)
7. Kirbas, C., Quek, F.: A review of vessel extraction techniques and algorithms, vol. 36, pp. 81–121. ACM, New York (2004)
8. Kolmogorov, V., Zabih, R.: What energy functions can be minimized via graph cuts, vol. 26, pp. 65–81. IEEE Computer Society, Los Alamitos (2004)
9. Lopez, A.M., Lumbrebras, F., Serrat, J., Villanueva, J.J.: Evaluation of methods for ridge and valley detection, vol. 21, pp. 327–335. IEEE Computer Society, Los Alamitos (1999)
10. Mirzaalian, H., Hamarneh, G.: Vessel scale-selection using mrf optimization. In: CVPR, pp. 3273–3279. IEEE Computer Society, Los Alamitos (June 2010)
11. Schaap, M., van Walsum, T., Niessen, W.: Standardized evaluation methodology and reference database for evaluating coronary artery centerline extraction algorithms, vol. 13, pp. 701–714. Elsevier, Amsterdam (2009)
12. Schneider, M., Sundar, H.: Automatic global vessel segmentation and catheter removal using local geometry information and vector field integration. In: IEEE ISBI, pp. 45–48. IEEE Press, Los Alamitos (2010)

Robust Physically-Constrained Modeling of the Mitral Valve and Subvalvular Apparatus

Ingmar Voigt^{1,2}, Tommaso Mansi¹, Razvan Ioan Ionasec¹,
Etienne Assoumou Mengue¹, Helene Houle³, Bogdan Georgescu¹,
Joachim Hornegger², and Dorin Comaniciu¹

¹ Image Analytics and Informatics, Siemens Corporate Research, Princeton, USA

² Pattern Recognition Lab, University of Erlangen-Nuremberg, Germany

³ Ultrasound, PLM-Innovation, Siemens Healthcare, Mountain View, CA, USA

Abstract. Mitral valve (MV) is often involved in cardiac diseases, with various pathological patterns that require a systemic view of the entire MV apparatus. Due to its complex shape and dynamics, patient-specific modeling of the MV constitutes a particular challenge. We propose a novel approach for personalized modeling of the dynamic MV and its subvalvular apparatus that ensures temporal consistency over the cardiac sequence and provides realistic deformations. The idea is to detect the anatomical MV components under constraints derived from the biomechanical properties of the leaflets. This is achieved by a robust two-step alternate algorithm that combines discriminative learning and leaflet biomechanics. Extensive evaluation on 200 transesophageal echochardiographic sequences showed an average Hausdorff error of 5.1mm at a speed of 9sec, which constitutes an improvement of up to 11.5% compared to purely data driven approaches. Clinical evaluation on 42 subjects showed, that the proposed fully-automatic approach could provide discriminant biomarkers to detect and quantify remodeling of annulus and leaflets in functional mitral regurgitation.

1 Introduction

Dysfunctional mitral valves (MV) are found in a large variety of cardiovascular diseases. Mitral valve prolapse for instance, where the leaflets billow into the left atrium due to dysfunctional chordae, is a common cause of regurgitation. A pathological left ventricular (LV) function, like in ischemic or idiopathic dilated cardiomyopathy, can also yield severe mitral valve insufficiency [1] due to an enlarged mitral annulus or abnormal positioning of the papillary muscles. Because the correct functioning of the MV relies on the tight synchronization of the LV, papillaries and leaflets dynamics, its evaluation requires a systemic view of the entire MV apparatus for accurate diagnosis, treatment planning and patient management. Echocardiography is the state-of-the-art modality for MV assessment. However, the complex appearance of the MV, the fast dynamics, the large morphological and functional variations among patients, and the varying image quality (low signal-to-noise ratio, signal dropout, time varying appearance) make the automatic modeling of the MV apparatus very challenging.

So far, existent valve models are generated manually or, at best, interactively [2]. The process is elaborate, tedious, and prone to inter-user variability. A detailed 4D model of the aorta and the MV, but without papillaries, was recently proposed [3]. In this work, the valves are estimated from the images using machine learning and statistical shape models. While these approaches have proven to be robust in many applications [4], their performance may decrease in the presence of image artifacts and in severe pathological morphologies. Furthermore, temporal consistency and physiologically realistic deformations are not guaranteed. In contrast, deformable models integrating prior knowledge about the cardiac deformation have been proposed to estimate the cardiac motion from images [5,6], with promising results. However, their application for valve dynamics is challenged by the very rapid valve motion. Furthermore, these methods are sensitive to initialization and local minima as their data terms rely on local image features.

Motivated by the above-mentioned limitations, we aim in this paper to estimate a personalized model of the mitral valve that is temporally and physiologically consistent. To that end, we propose a novel discriminative learning technique that is constrained by a biomechanical model of the MV leaflets. The model is automatically initialized in the images and local minima are avoided through incremental search and discriminative learning. The biomechanical constraint is ensured by solving a dynamic system between time frames. Besides, our model includes the MV annulus, anterior and posterior leaflets, and the subvalvular apparatus represented by landmarks at the papillary tips. To the best of our knowledge this is the first time that machine learning methods are integrated with biomechanical constraints. Extensive evaluation on 200 transesophageal echocardiographic (TEE) sequences show an average Hausdorff error of 5.1mm at a speed of 9sec per volume, which constitutes an improvement of up to 11.5% compared to state-of-the-art approaches [3,7], which are purely data driven. A clinical evaluation on 42 subjects is carried out, showing that our method can provide reliable discriminant biomarkers to detect and quantify remodeling of MV shape and dynamics.

2 Methods

2.1 General Overview of the Algorithm

Our goal is to estimate a physiologically realistic 4D model of the MV and its subvalvular apparatus from 3D+t cardiac images. The idea is to constrain a machine learning approach such that the estimated point distribution model $S \in \Omega$, Ω being the space of physiological surface shapes, satisfies MV biomechanics. Although the proposed approach is applied on the entire 4D time series, in the following we describe the algorithm between two frames I_{t_1} and I_{t_2} , at time t_1 and t_2 only, for the sake of clarity. The constrained problem is written as:

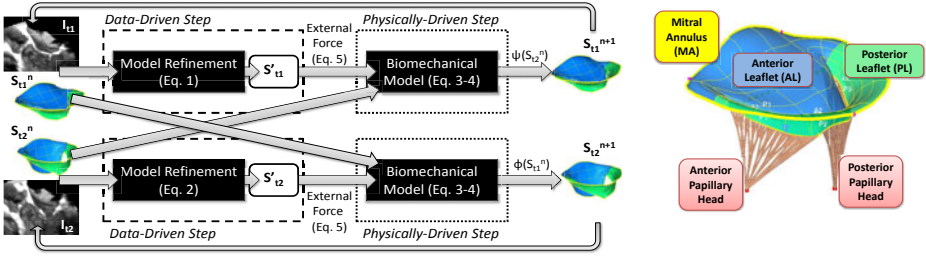


Fig. 1. *Left panel:* Proposed framework for physically-constrained MV estimation. See text for details. *Right panel:* Mitral valve apparatus model in systole.

$$\begin{cases} S_{t_1} = \arg \max_{\hat{S} \in \Omega} p(\hat{S} | I_{t_1}) & (1) \\ S_{t_2} = \arg \max_{\hat{S} \in \Omega} p(\hat{S} | I_{t_2}) & (2) \\ S_{t_2} = \phi(S_{t_1}) \text{ and } S_{t_1} = \psi(S_{t_2}) & (3) \end{cases}$$

where p is a discriminative probabilistic model and ϕ and ψ are the physiological deformations calculated by solving a dynamic system. To achieve computational efficiency, we solve this high-dimensional constrained and coupled problem with a two-step optimization procedure, as illustrated in Fig. 1. At a given iteration n of the algorithm loop, let $S_{t_1}^n$ and $S_{t_2}^n$ be the current physically constrained models in I_{t_1} and I_{t_2} respectively. In the first “data-driven” step, we refine $S_{t_1}^n$ using discriminative machine learning on I_{t_1} to obtain a new estimate S'_{t_1} (Eq. 1). In the second “physically-driven” step, we deform $S_{t_2}^n$ according to a biomechanical model ψ with external force calculated from S'_{t_1} to obtain an updated physically-constrained $S_{t_1}^{n+1}$ (Eq. 3). Thus, the updated model in image I_{t_1} is a deformed realization of $S_{t_2}^n$ from image I_{t_2} subject to data-driven information related to I_{t_1} . $S_{t_2}^{n+1}$ is computed in parallel following the same procedure. These two-steps, further detailed below, are iterated until convergence.

2.2 Data-Driven Step: Estimation of Mitral Valve Apparatus Model

Anatomical Model Parameters. The anatomical model S of the MV and its subvalvular component comprises: the mitral annulus (MA), the anterior and posterior leaflets and the anterior and posterior papillary heads (Fig. 1 right). More precisely, S is a point distribution model consisting of 986 points and 1792 triangles with consistent parameterization derived from anatomical landmarks (three trigones, three commissures, two leaflet tips and two papillary heads) [3].

Learning-Based Spatial Estimation. The parameters of the anatomical model S are estimated from the images using a hierarchical discriminative learning algorithm as in [3] (Eq. 1 and Eq. 2). The conditional probability $p(S|I)$ is incrementally modeled within the Marginal Space Learning (MSL) framework,

which consists in training and detecting on marginal spaces using the Probabilistic Boosting Tree (PBT) with Haar and steerable features [7]. Different detectors are employed for the mitral annulus and free-edges contours and the leaflet surfaces to improve detection accuracy.

2.3 Physically-Driven Step: Biomechanical Constraint

We now constrain the estimations in I_{t_1} and I_{t_2} according to a biomechanical model of the MV leaflets. Let $S_{t_1}^n$ and $S_{t_2}^n$ be the current physically-constrained model and S'_{t_1} and S'_{t_2} the new estimates computed as described in the previous sections. The updated constrained model $S_{t_1}^{n+1}$ related to I_{t_1} is obtained by deforming $S_{t_2}^n$ towards S'_{t_1} . This deformation, ψ in Eq. 3, is achieved by solving the dynamic system

$$M\ddot{\mathbf{U}} + C\dot{\mathbf{U}} + \mathbf{K}\mathbf{U} = \mathbf{F}_{ext} \quad (4)$$

\mathbf{U} is the displacement vector of the vertices of $S_{t_2}^n$, $\dot{\mathbf{U}}$ their velocity and $\ddot{\mathbf{U}}$ their acceleration. M is a diagonal mass matrix (leaflet mass density $\rho = 1.04 \text{ g/mL}$), C is a Rayleigh damping matrix, $C = 0.1(M + \mathbf{K})$, and \mathbf{K} is the stiffness matrix of the internal elastic forces. We approximate the tissue properties of the leaflets by a linear isotropic material to optimize computational efficiency for fast estimation. The leaflet thickness is set to 2mm, the average thickness observed in our datasets. Near-incompressibility is achieved with a Poisson ratio ν of 0.488 and a Young modulus E of 6.2 MPa is used [8].

\mathbf{F}_{ext} is the external force that drives $S_{t_2}^n$ towards the new estimate S'_{t_1} . We want the result to be as close as possible to S'_{t_1} but that preserves the tangential motion generated by the internal forces. We therefore move the vertices $\mathbf{v}_{t_2}^n$ of $S_{t_2}^n$ along their normal direction \mathbf{n} , towards their corresponding vertex \mathbf{v}'_{t_1} in S'_{t_1} . Furthermore, we weight that force according to the uncertainty in the data term $p(\mathbf{v}'_{t_1} | I_{t_1})$ such that positions with low confidences have little influence on the leaflet deformation, while high confidences result in high influence. Thus, \mathbf{F}_{ext} is written as:

$$\mathbf{F}_{ext}(\mathbf{v}_{t_2}^n) = -\kappa p(\mathbf{v}'_{t_1} | I_{t_1})(\mathbf{v}'_{t_1} - \mathbf{v}_{t_2}^n) \cdot \mathbf{n} \quad (5)$$

κ is a weight parameter, which in our experiments we empirically set to 0.1. The vertices $\mathbf{v}_{t_2}^n$, and thus the force \mathbf{F}_{ext} , are updated at every time step of the resolution of the dynamic system in Eq. 4, which is solved using co-rotational triangular finite element methods (FEM) to cope with large deformations and rotations [9]. An implicit Euler solver is employed to update mesh positions. The deformation ends when the average relative displacement of the surface vertices is lower than the image resolution (typically 1 mm).

Intuitively, the final deformed model $\psi(S_{t_2}^n) = S_{t_1}^{n+1}$ matches the estimated model S'_{t_1} . However, here, the new model $S_{t_1}^{n+1}$ takes into account the internal forces and is robust to outliers with low confidence from the data term. It is also physically consistent with $S_{t_2}^n$. We apply the same procedure, symmetrically, to estimate $S_{t_2}^{n+1}$, and iterate. The algorithm is extended to 4D sequences by propagating the models over the cardiac sequence using a cyclic constraint and

Table 1. Estimation accuracy in terms of maximum (Hausdorff) and average (point-to-mesh) distances in mm to the ground truth for the biomechanically constrained and unconstrained, purely data driven approaches. Notable improvements are highlighted in bold faced font for corresponding metrics for both constrained and unconstrained estimation results.

	Max error (Hausdorff)						Average error (Point-to-mesh)					
	constrained			unconstrained			constrained			unconstrained		
Anatomy	mean	std	90%	mean	std	90%	mean	std	90%	mean	std	90%
Annulus	4.41	1.60	5.98	4.57	1.79	6.81	1.88	0.57	2.54	1.81	0.58	2.46
Free Edge	5.32	1.85	7.53	5.73	2.69	8.50	2.44	0.70	3.32	2.39	0.85	3.26
Leaflets	5.57	1.68	7.51	5.70	2.49	8.36	3.11	1.02	4.14	3.02	1.11	4.36

iterating until convergence. The algorithm ends when the norm of the maximum update displacement in all the frames is below the image resolution. In practice, the proposed algorithm converged fairly fast. This may be explained by the fact that each surface is deformed using the other estimated model. As the model comes closer to the maximum PBT, an equilibrium is found.

3 Experimental Results

The evaluation of the proposed method splits into two subparts: *i*) the quantitative evaluation of the model estimation performance and *ii*) a clinical evaluation in terms of a predictive biomarker for diagnosis and therapy planning.

Estimation performance. Our method has been evaluated using a comprehensive set of 200 4D TEE studies from 120 patients, which were acquired using different capture ranges and image resolutions. Each study was associated with ground-truth models, which were manually fitted to the image data. The performance results were obtained using three-fold cross validation experiments for both biomechanically constrained and unconstrained estimation schemes. The method performs with a speed of 9 seconds per volume on a standard desktop machine (Intel Core Duo 2.66 GHz, 2 GB RAM). The biomechanical deformations required 5 to 25 steps to converge. The overall algorithm converged in 5 to 7 iterations and no convergence problems could be observed. The average point-to-mesh distance and Hausdorff distance are reported in Table 1.

While the average point-to-mesh accuracy is comparable for both methods – considering the 1mm image resolution –, the Hausdorff distances underlines the specific strength of the biomechanical constraint by providing improved robustness over the purely data driven approach esp. in cases with challenging image quality and constitutes an improvement of up to 11.5%, while at the same time preserving a reasonable estimation speed of 9 seconds per volume. Particularly the leaflets and free edges of the valve, which are governed by rapid motion and signal dropout - esp. in diastole - benefit from the constrained estimation scheme as indicated by a reduced mean, standard deviation and 90-percentile. Fig. 2 illustrates a specific case for which the unconstrained estimation algorithm fails

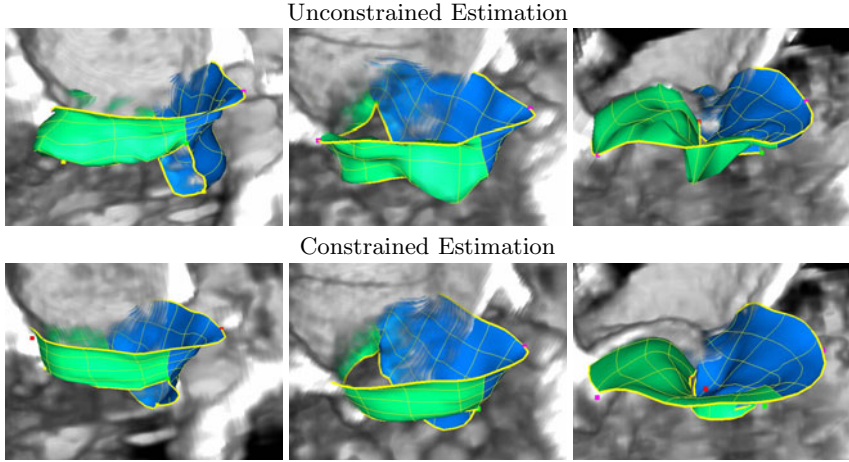


Fig. 2. Exemplary estimation results on three patients. While the unconstrained estimation alone is subject to image noise and signal dropout, the biomechanical constraint is compensating for such outliers and provides realistic shapes and smooth surfaces.

to delineate the mitral valve due to image noise, while the biomechanical constraint compensates for such outliers and provides realistic shapes and smooth surfaces. As the proposed algorithm utilizes a biomechanical model of the valves, temporal consistency and physiological variation in anatomical dimensions respects the near-incompressibility of the tissue. The relative temporal variation in surface area and leaflet length is consistently lower using the biomechanical constraint and quantifies to 10% vs. 14% for the constrained vs. unconstrained estimation.

Clinical Evaluation. The contributions of mitral annular (MA) and leaflet (ML) remodeling to functional mitral regurgitation (FMR) is a key factor in the consideration of mitral valve annuloplasty. Selection of the synthetic ring size and type is based on the quantification of MA and ML remodeling. We thus tested the feasibility of the proposed MV modeling approach to quantify MA and ML remodeling in FMR [10].

The mitral valve was estimated from 15 healthy subjects and 27 patients with moderate or severe FMR. Among the patients, 12 presented with normal annulus size (Group 1) and 15 with a dilated annulus (Group 2). From the 4D anatomical model of the MV, we calculated different metrics reported in Table 2. The dynamic change of the antero-posterior (AP) annular diameter, computed as the difference diameter between early systole and early diastole, significantly discriminated the normals from FMR Group 1 (Fig. 3). In particular, it indicates a reduced accentuation of the saddle shape with consequent reduction in leaflet coaptation. In addition to a dilated, akinetic annulus, the anterior surface length was markedly increased in FMR Group 2, which characterizes significant ML remodeling. The choice of ring size and type can be made based on these

Table 2. Model-based measurements for quantification of functional mitral regurgitation (FMR) and surgical planning: Mean and standard deviation in respective groups reported in mm.

Dimension	healthy	FMR Group 1	FMR Group 2
Annular AP diameter	29.2 ± 3	33.1 ± 1	34.4 ± 4
Anterior leaflet height	20.9 ± 3	23.9 ± 4	26.9 ± 5
Annular circumference	112.1 ± 9	114.4 ± 5	129.0 ± 12
Dynamic AP diameter change			
Early diastole	30.6 ± 4	33.3 ± 2	33.9 ± 4
Early systole	26.1 ± 3	32.9 ± 3	32.5 ± 4

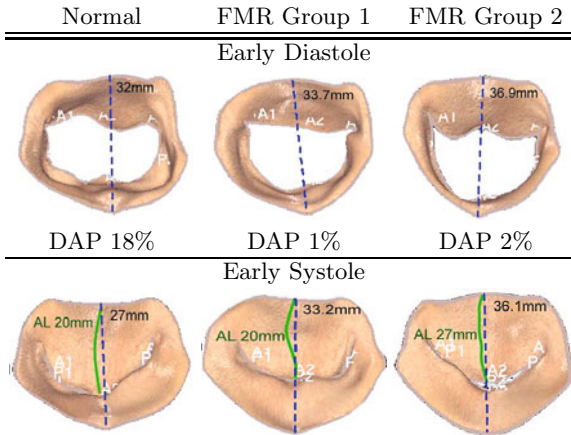


Fig. 3. Dynamic AP (DAP) Diameter (blue dashed line) varies by 18% over the cardiac cycle in healthy subjects in contrast to FMR cases (DAP = 1-2%). Anterior leaflet height (green solid line) was normal in group 1 but not in group 2, indicating predominant annular remodeling in that population [10]. Reproduced with permission of the authors.

automated measurements. We can thus conclude that automated 3-D quantitative surgical anatomy in FMR 1) shows that an akinetic annulus is an early basis for MR, followed by annular dilatation and anterior leaflet lengthening, and 2) that these mechanistic insights and the quantitative characterization of the pathological anatomy can aid surgical decision-making.

4 Conclusion

We proposed a novel approach for physiologically and temporally compliant estimation of the mitral valve by combining discriminative learning methods with biomechanical constraints. The data-driven step ensures efficiency and robustness while the physically-driven step guarantees physiological consistency. Our

method is fully automatic, robust to initialization and image artifacts. Furthermore, our mitral valve model comprises the mitral annulus, anterior and posterior leaflets and for the first time, the subvalvular apparatus. Quantitative evaluation on 200 datasets demonstrated that our constrained method outperforms purely data driven approaches by 11.5% in terms of model estimation (average Hausdorff error of 5.1mm). Clinical evaluation demonstrated that the proposed approach could provide reliable biomarkers for mitral annulus and leaflets remodeling in functional mitral regurgitations. Future work will further investigate effects of the different degrees of freedom such as anisotropic and non-linear material properties.

References

1. Borger, M., Alam, A., Murphy, P., Doenst, T., David, T.: Chronic ischemic mitral regurgitation: Repair, replace or rethink? *Ann. Thorac. Surg.* (81), 1153–1161 (2006)
2. 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. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 366(1879), 3411 (2008)
3. 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 4-D Cardiac CT and TEE. *IEEE Transactions on Medical Imaging* 29(9), 1636–1651 (2010)
4. Heimann, T., Meinzer, H.P.: Statistical shape models for 3d medical image segmentation: A review. *Medical Image Analysis* 13(4), 543–563 (2009)
5. Papademetris, X., Sinusas, A.J., Dione, D.P., Duncan, J.S.: Estimation of 3D left ventricular deformation from echocardiography. *Med. Image Anal.* 5, 17–28 (2001)
6. Sermesant, M., Delingette, H., Ayache, N.: An electromechanical model of the heart for image analysis and simulation. *IEEE Transactions in Medical Imaging* 5(25), 612–625 (2006)
7. Zheng, Y., Barbu, A., Georgescu, B., Scheuering, M., Comaniciu, D.: Four-chamber heart modeling and automatic segmentation for 3-d cardiac ct volumes using marginal space learning and steerable features. *IEEE Transactions on Medical Imaging* 27(11), 1668–1681 (2008)
8. 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. *Journal of Heart Valve Disease* 18(1), 28–34 (2009)
9. Nesme, M., Payan, Y., Faure, F.: Efficient, physically plausible finite elements. *Eurographics (short papers)*, 77–80 (2005)
10. Calleja, A., Stiver, K., Thavendiranathan, P., Liu, S., Ionasec, R., Voigt, I., Houle, H., Michelis, N.D., Ryan, T., Vannan, M.: Automated Quantitative 3-D Echocardiography of The Surgical Mitral Valve Anatomy in Functional Mitral Regurgitation to Guide Mitral Valve Repair. In: *Proceedings of the 22nd Annual Scientific Sessions of the American Society of Echocardiography* (2011)

Regurgitation Quantification Using 3D PISA in Volume Echocardiography

Leo Grady¹, Saurabh Datta², Oliver Kutter¹, Christophe Duong²,
Wolfgang Wein³, Stephen H. Little⁴, Stephen R. Igo⁴,
Shizhen Liu⁵, and Mani Vannan⁵

¹ Siemens Corporate Research, Princeton, NJ, USA

² Siemens Healthcare, Ultrasound Business Unit, Mountain View, CA, USA

³ White Lion Technologies AG, Munich, Germany

⁴ Methodist DeBakey Heart and Vascular Center, Houston, TX, USA

⁵ Ohio State University, Columbus OH, USA

Abstract. We present the first system for measurement of proximal isovelocity surface area (PISA) on a 3D ultrasound acquisition using modified ultrasound hardware, volumetric image segmentation and a simple efficient workflow. Accurate measurement of the PISA in 3D flow through a valve is an emerging method for quantitatively assessing cardiac valve regurgitation and function. Current state of the art protocols for assessing regurgitant flow require laborious and time consuming user interaction with the data, where a precise execution is crucial for an accurate diagnosis. We propose a new improved 3D PISA workflow that is initialized interactively with two points, followed by fully automatic segmentation of the valve annulus and isovelocity surface area computation. Our system is first validated against several *in vitro* phantoms to verify the calculations of surface area, orifice area and regurgitant flow. Finally, we use our system to compare orifice area calculations obtained from *in vivo* patient imaging measurements to an independent measurement and then use our system to successfully classify patients into *mild-moderate regurgitation* and *moderate-severe regurgitation* categories.

1 Introduction

Valvular heart diseases are a major cause of death in developed countries, which are estimated to affect 2.5% of the population and are the underlying cause of over 43700 deaths and 93000 hospital discharges in 2006 in the US. Pooled data from multiple studies suggest that the prevalence increases with age from 0.7% in participants 18–44 years of age to 13.3% in participants ≥ 75 years of age. The cause of these diseases can be congenital or acquired and, depending on the severity, treatment might involve medication, surgical valve repair or replacement. Timely diagnosis and accurate assessment of regurgitation severity is important for appropriate clinical decision making and optimal patient outcome [3].

Doppler ultrasound is routinely used to non-invasively evaluate intracardiac blood flow and detect regurgitant lesions for all cardiac valves. Several indexes

have been developed to assess severity of regurgitation using color Doppler, Pulsed Wave (PW) and Continuous Wave (CW) Doppler. However, the use of color Doppler is limited to a semi-quantitative assessment of regurgitant jets directly or indirectly and is often dependent on user expertise. One direct, but laborious, method for quantifying regurgitation is the measurement of vena contracta (VC), which is the narrowest cross-sectional area of the jet. Although VC measurement is a close approximation of the regurgitant orifice, this measure has limited accuracy due to dependence on system resolution, gain setting, acquisition view and user expertise [1].

The proximal isovelocity surface area (PISA) is a promising method for quantifying valve regurgitation. The PISA measurement is derived from the fluid dynamics principle that blood velocity increases as it approaches the valve orifice, forming concentric hemispheric shells of increasing velocity and decreasing surface area. Using appropriate settings of color Doppler imaging, these hemispheres can be visualized. Clinically, this estimation requires geometric assumptions and multiple steps. The effective regurgitant orifice area (EROA) or instantaneous flow rate (mL/s) can be computed from PISA via

$$\text{EROA} = \frac{\text{PISA} \times \text{Aliasing Velocity}}{\text{Peak Velocity}}. \quad (1)$$

The EROA is an important quantity to determine the best course of patient treatment [2]. Additionally, a PISA measurement makes it possible to directly calculate the regurgitation volume via

$$\text{Regurgitation volume} = \int \text{PISA} \times \text{Aliasing Velocity} dt. \quad (2)$$

In 2D ultrasound acquisitions, the EROA is computed with a hemispheric geometric assumptions. Recently, 3D volume color Doppler acquisitions have made it possible to utilize more appropriate geometric assumptions [7,11] and provide better accuracy. However, these studies required a prolonged workflow involving multiple steps and computations, creating a barrier for clinical acceptance. Additionally, the volumetric color flow quantification accuracy in these studies is still dependent on system parameters such as acquisition (e.g., gated, stitched), angle, temporal and spatial resolution, system settings and workflow [5]. The clinical acceptance of the PISA technique will require real-time volume color Doppler capability on the acquisition system, an enhanced workflow, accurate quantification and elimination of geometric approximation. Fig. 1 illustrates the PISA and VC calculations from a color Doppler acquisition.

In this work, we use the capabilities of a state of the art real-time 3D cardiac ultrasound system in conjunction with 3D image segmentation techniques in order to improve the accuracy of the PISA measurement and simplify the workflow. Section 2 first explains our volumetric image acquisition hardware, and the segmentation algorithm used to extract the mitral annulus and the PISA surface. Validation on a range of *in vitro* and *in vivo* acquisitions are presented in Section 3, followed by a discussion in Section 4.

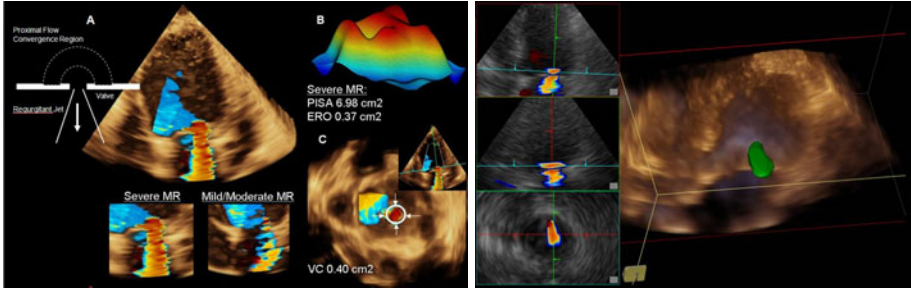


Fig. 1. Left: (A) Illustration of different cardiac mitral regurgitation (MR) and proximal isovelocity surface area (PISA) methods, (B) PISA surface generated from 3D manual measurement, (C) Manual measurement of vena contracta (VC). Right: An example PISA surface obtained from our system (different dataset).

2 Methods

Our system for accurate PISA quantification required hardware capability to acquire full volume color Doppler data along with software improvements. The hardware allows for sufficiently high volume rate acquisition and the software improvements were necessary to achieve an accurate and robust quantification. Our system was interactive, asking the user to place a point at the valve annulus and at the coaptation point (in the apical 4 or 2 chamber view).

2.1 Echocardiography Imaging and Visualization

We used a 2D matrix phased array ultrasound system (the ACUSON SC2000TM, 4Z1c transducer, which is now commercially available) for dual B-mode and color Doppler volume imaging. High volume rate acquisition is critical for an accurate quantification of the PISA surface volume, because low acquisition rates cause velocity discrepancies between multiple acquisition planes. These discrepancies create an inaccurate velocity surface estimation. The volume acquisition at higher rates is achieved by decoupling the transmit foci in the azimuthal and elevational directions. In azimuth, a plane-wave was used and in elevation, a virtual point source was used. For adequate spatial sampling of the volume in receive mode, we used a large number of parallel receive lines (40 or 64) for every transmit line. By combining a weakly focused transmit with many parallel receive lines, a high frame rate was achieved without a loss of spatial resolution. For example, a volume rate of $\approx 30/s$ is possible with fields of view of $70 \times 70^\circ$ in B-mode and $30 \times 30^\circ$ in color Doppler. Such volume rates are both comparable to conventional 2D color Doppler frame rates, and adequate for accurate PISA estimation. The exact imaging resolution depends on depth, but it was $\approx 1\text{mm}$ for the Doppler acquisition and sub-mm for the B-mode acquisition. The imaging results in Fig. 1 were acquired using our system.

2.2 Localization of the Valve Annulus

The extent of the valve annulus must be known in order to both quantify the flow through the valve and to provide a reference location from which to calculate the PISA surface. Our method can be used to localize the valve annulus for any target valve which has been specified by a single point inside the valve annulus (which we call the *valve point*).

Segmentation of the valve annulus is a challenging problem, since the valve is essentially a hole joining two chambers, creating poorly-defined boundaries. Additionally, the shape of the annulus and the surrounding anatomy varies depending on the target valve, patient and cardiac phase. We selected the random walker algorithm [4] for this segmentation task since it is known to behave well with poorly defined (or absent) boundaries and it makes no explicit assumptions on the annulus shape (which helps make the segmentation robust to large variability). Our system performs an interactive segmentation of the annulus, in contrast to [9] which employs an automatic, but more complex system.

Our strategy for utilizing the random walker algorithm for annulus segmentation is to automatically identify some voxels inside/outside the annulus which can be used to constrain the segmentation algorithm. In order to set seeds for the random walker algorithm, we adopt a simple model of the annulus (assuming an open valve) as a hypoechoic connection between two hypoechoic chambers. In order to determine the appropriate inputs to the random walker, we first determine the model orientation and then establish appropriate seeds. To improve the speed of the calculations, the annulus segmentation was limited to a 5cm 3D ROI surrounding the valve point. For purposes of the model, we define a voxel to be *hypoechoic* if it falls below the median intensity in the ROI.

The orientation vector of the annulus model is determined by calculating the largest principal component of the hypoechoic voxels. Given the valve point and the orientation, we may identify annulus voxels (foreground seeds) for the random walker inside the annulus plane. The annulus plane is determined to be the hypoechoic part of the orthogonal plane connected to the valve point. In order to restrict the annulus segmentation, background seeds are placed parallel to the annulus plane on both sides at a distance of 2.5cm (this value is used only to approximately set seeds and therefore does not need to be precise). With these input seeds, the random walker segmentation algorithm is employed to produce the annulus segmentation on the B-mode image. This approach to valve annulus segmentation may be applied to any of the cardiac valves (given a valve point), although PISA measurements are typically used to assess regurgitation at the mitral valve and left ventricular outflow tract (LVOT).

2.3 PISA Measurement

Once the location of the jet and valve coaptation point is located by the user, the PISA calculation may be performed. The identification of PISA in the 3D space allows for the computation of regurgitant flow during the cardiac phase of interest using (2). Our system expects that the desired isovelocity value and the

direction of interest (relative to the annulus or valve coaptation line) is specified by the user and that the target isovelocity surface is contained entirely within the acquisition. Our strategy for PISA measurement is to generate a segmentation of the isovelocity region in the color Doppler data from which a surface can be segmented and a surface area may be calculated. Since the PISA is calculated above the coaptation point (relative to the valve annulus), the flows below the coaptation point (in the direction of the annulus) are ignored.

The isovelocity region above the coaptation point (i.e., in the direction away from the annulus) can be challenging to segment as a result of flow aliasing or confounding flow through a second valve (e.g., measuring PISA at the mitral valve may be confounded by the nearby flow through the LVOT). To account for these two challenges, we again use the random walker but this time we employ a novel *directed* graph which is designed to prevent cojoinment with a secondary flow. Specifically, for a target isovelocity value of γ , we establish the directed graph weight between two voxels v_i and v_j as $w_{ij} = 1$ if $f_i, f_j > \gamma$ and $f_i < f_j$, or if $f_i, f_j \leq \gamma$, where f_i represents the magnitude of the flow velocity at voxel v_i . Otherwise, $w_{ij} = \epsilon$. Optimization of the random walker with a directed graph is detailed in [10]. The coaptation point is treated as a foreground seed and the volume borders are treated as background seeds in the PISA segmentation.

The voxel-based segmentation result is smoothed using a 3D Gaussian kernel, and an isosurface mesh is successively computed using the Marching Cubes algorithm. The intersection with the mitral annulus segmentation is then removed from the mesh, yielding the hemispherical isosurface. Finally, all mesh vertex locations are transformed from acoustic to Cartesian space for computing the actual 3D PISA surface area measurement.

3 Experiments and Results

Our validation was designed to address the following issues: 1) Accuracy of the surface area computation, 2) Accuracy of the EROA calculation for various orifice shapes, relative to ground truth and relative to standard EROA approximation methods, 3) Accuracy of the regurgitant volume estimation, relative to ground truth and standard approximation methods, 4) Accuracy of EROA calculation on clinical patients, 5) The ability of our system to distinguish patients with mild-moderate regurgitation from patients with moderate-severe regurgitation. The first three issues were evaluated using a precise *in vitro* phantom (shown in Fig. 2) which could be controlled to produce calibrated flows and regurgitations through synthetic orifices of different shapes. The last two issues were probed using a retrospective study on a clinical patient population.

PISA segmentation and surface area computation was evaluated with a calibrated phantom. A hollow sphere with known surface area was imaged with the full volume 3D ultrasound system in the clinical configuration (shown in Fig. 2). The purpose of this experiment was to test the resolution limit and geometric consistency of the acquisition as a function of depth and beam overlap. Compared to the reference surface area of 49.09cm², the mean surface area computed by our system was 49.87cm² over five trials of acquisition/segmentation.

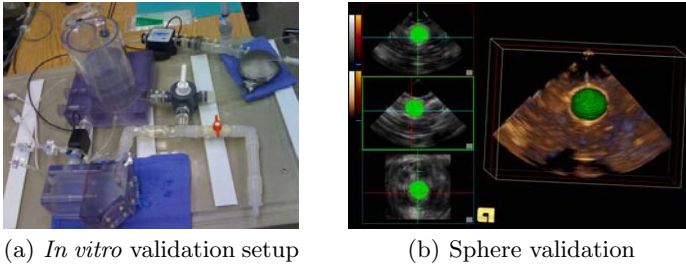


Fig. 2. Surface area validation. (a) Calibrated *in vitro* phantom. (b) Validation of our PISA quantification system for measuring the surface area of a sphere with known diameter.

To probe the accuracy of the EROA measurement, we created an *in vitro* phantom (described in [6]) with several orifice shapes in which the flow could be controlled to lie within a physiologically relevant severity of regurgitation (12–57mL/beat). The flow loop was constructed to create physiologically relevant pressure differential and flow across the orifice of a regurgitant valve [6]. Data acquisition was performed using a standard clinical 3D full volume ultrasound scanner. Our method for measuring PISA was used to estimate the physiological flow convergence orifice and multiple frames were integrated to compute the flow for the entire cycle. A flow meter was placed in line to measure the true flow in the closed loop to measure regurgitation volume in the regurgitation flow loop. The values of effective regurgitant orifice area (EROA) and regurgitant volume were computed from the PISA measurements for 60 trials. The EROA measurement was compared with the true anatomic orifice area and the regurgitant volume was compared with the flow meter values for 32 trials. Additionally, we compared the same EROA measurements to the common clinical practice of hemispheric PISA estimation. Our system for PISA quantification consistently produced better EROA estimates than the reference method, as shown in Fig. 3. Note that the reference line in Fig. 3 refers to the value of the true orifice area. However, the *physiologic orifice* measured by the EROA is known to be 80–90% smaller than the true orifice area [2]. We also computed the regurgitant volume from our system and compared this volume by the in line flow meter measurement used as the ground truth for regurgitant volume in our system (shown in Fig. 3). The conventional hemispheric method severely underestimated the EROA in all orifices and flow conditions including the circular orifice cases as shown in Fig. 3. Our method for measuring PISA performed better for all orifice shapes and flow conditions. The computation of regurgitant flow volumes also reflect a severe underestimation by the conventional PISA method when compared with the true regurgitant volume measured by a flow meter. In contrast, our PISA quantification system performed better in all cases irrespective of orifice geometry or flow condition as can be observed in Fig. 3. The *in vitro* work demonstrates the improvement in estimation of regurgitation using our method for measuring PISA in a controlled experimental setting.

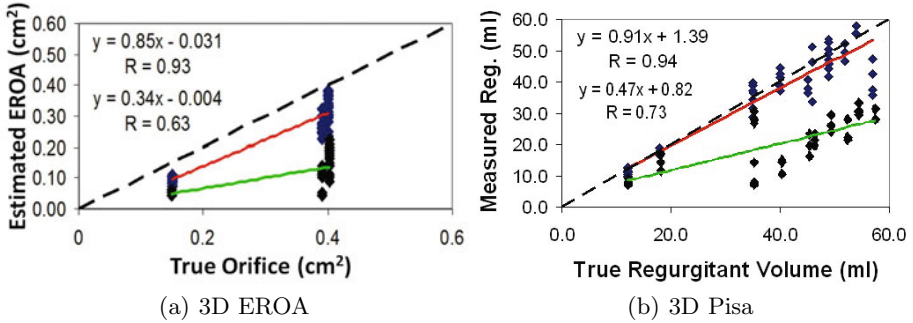


Fig. 3. Results of experimental validation using an *in vitro* phantom to measure EROA and regurgitation volume. Bottom/green: Conventional spherical estimate. Top/red: System estimate. The spherical method severely underestimated the EROA in all orifices and flow conditions including the circular orifice (similar observations were made in [8]). Our method for measuring PISA performed better for all orifice shapes and flow conditions. The line in Fig. 3 refers to the value of the true orifice area, although the *physiologic orifice* measured by the EROA is known to be 80–90% smaller than the true orifice area [2].

In order to determine the accuracy of our system on patient data, we compared the EROA measurements obtained by our system with the measurement of vena contracta. Vena contracta measurements were performed manually by a clinical expert by tracing the flow convergence in a carefully positioned and aligned plane. The expert vena contracta measurements were compared with the EROA measurements obtained from our system. For a patient sample size of 33, we observed a statistically significant correlation ($R=0.61$, $p=0.002$) between the EROA measurement and vena contracta measurements.

As a final test of our system on a clinical patient sample, we retrospectively examined a group of 35 patients which were segregated according to the clinical diagnosis of mitral regurgitation being mild-moderate or moderate-severe. Using our system, we measured EROA values from all patients in both groups. We determined that the mean EROA value for the mild-moderate regurgitation patient group was $0.24 \pm 0.07 \text{ cm}^2$ and the mean EROA value for the patient group with moderate-severe regurgitation was $0.43 \pm 0.13 \text{ cm}^2$. These measurements were statistically distinguishable ($p < 0.001$), which indicated that our system can be used to distinguish these two patient populations.

4 Discussion

We have presented a system and workflow for fast acquisition and quantification of PISA, leading to accurate calculations of EROA and regurgitation volume. Our system utilizes 3D B-mode and Doppler flow volumetric echocardiography sequence data. To our knowledge, this is the first time that the complex 3D PISA diagnostic evaluation workflow was addressed by an efficient computational sys-

tem. Extensive *in vitro* experiments establish that our system accurately measures surface area, EROA and regurgitant flow, especially as compared to the calculations obtained from the conventional spherical approximation. Further *in vivo* experiments with a patient population demonstrate that our system effectively measures EROA when cross-validated with a manual vena contracta measurement and that our system makes it possible to accurately distinguish patients with mild-moderate regurgitant flow from patients with moderate-severe regurgitant flow. In future work, the interactive specification of the valve and coaptation points may be replaced by an automatic localization.

References

1. Buck, T., Plicht, B., Hunold, P., Mucci, R., Erbel, R., Levine, R.: Broad-beam spectral Doppler sonification of the vena contracta using matrix-array technology—a new solution for semi-automated quantification of mitral regurgitant flow volume and orifice area. *J. Am. Coll. Cardiol.* 45, 770–779 (2005)
2. DeGroff, C., Shandas, R., Valdes-Cruz, L.: Analysis of the effect of flow rate on the Doppler continuity equation for stenotic orifice area calculations: a numerical study. *Circ.* 97(16) (1998)
3. Enriquez-Sarano, M.: Timing of mitral valve surgery. *Heart*, 79–85 (2002)
4. Grady, L.: Random walks for image segmentation. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 28(11), 1768–1783 (2006)
5. Lang, R., Mor-Avi, V., Sugeng, L., Nieman, P., Sahn, D.: Three-dimensional echocardiography: The benefits of the additional dimension. *J. Am. Coll. Cardiol.* 48, 2053–2669 (2006)
6. Little, S., Igo, S., McCulloch, M., Hartley, C., Nosé, Y., Zoghbi, W.: Three-dimensional ultrasound imaging model of mitral valve regurgitation: design and evaluation. *Ultrasound in Med. Biol.* 34(4), 647–654 (2008)
7. Matsumura, Y., Fukuda, S., Tran, H., Greenberg, N., Agler, D., Wada, N., Toyono, M., Thomas, J.: Geometry of the proximal isovelocity surface area in mitral regurgitation by 3-dimensional color Doppler echocardiography: difference between functional mitral regurgitation and prolapse regurgitation. *AHJ*, 155–231 (2008)
8. Pirat, B., Little, S., Igo, S., McCulloch, M., Nosé, Y., Hartley, C., Zoghbi, W.: Direct measurement of proximal isovelocity surface area by real-time three-dimensional color Doppler for quantitation of aortic regurgitant volume: an *in vitro* validation. *J. Am. Soc. Echo.* 22(3), 306–313 (2009)
9. Schneider, R., Perrin, D., Vasilyev, N., Marx, G., del Nido, P., Howe, R.: Mitral annulus segmentation from 3D ultrasound using graph cuts. *IEEE TMI* 29(9), 1676–1687 (2010)
10. Singaraju, D., Grady, L., Vidal, R.: Interactive image segmentation of quadratic energies on directed graphs. In: *Proc. of CVPR* (June 2008)
11. Yosefy, C., Levine, R., Solis, J., Vaturi, M., Handschumacher, M., Hung, J.: Proximal flow convergence region as assessed by real-time 3-dimensional echocardiography: challenging the hemispheric assumption. *J. Am. Soc. Echo.* (2007)
12. Zoghbi, W., Enriquez-Sarano, M., Foster, E., Grayburn, P., Kraft, C., Levine, R., Nihoyannopoulos, P., Otto, C., Quinones, M., Rakowski, H.: Recommendations for evaluation of the severity of native valvular regurgitation with two-dimensional and Doppler echocardiography. *J. Am. Soc. Echo.* 16, 777–802 (2003)

Patient-Specific Mitral Leaflet Segmentation from 4D Ultrasound

Robert J. Schneider¹, Neil A. Tenenholtz¹, Douglas P. Perrin^{1,2},
Gerald R. Marx³, Pedro J. del Nido², and Robert D. Howe¹

¹ Harvard School of Engineering and Applied Sciences, Cambridge, MA, USA

² Department of Cardiac Surgery, Children's Hospital, Boston, MA, USA

³ Department of Cardiology, Children's Hospital, Boston, MA, USA

Abstract. Segmenting the mitral valve during closure and throughout a cardiac cycle from four dimensional ultrasound (4DUS) is important for creation and validation of mechanical models and for improved visualization and understanding of mitral valve behavior. Current methods of segmenting the valve from 4DUS either require extensive user interaction and initialization, do not maintain the valve geometry across a cardiac cycle, or are incapable of producing a detailed coaptation line and surface. We present a method of segmenting the mitral valve annulus and leaflets from 4DUS such that a detailed, patient-specific annulus and leaflets are tracked throughout mitral valve closure, resulting in a detailed coaptation region. The method requires only the selection of two frames from a sequence indicating the start and end of valve closure and a single point near a closed valve. The annulus and leaflets are first found through direct segmentation in the appropriate frames and then by tracking the known geometry to the remaining frames. We compared the automatically segmented meshes to expert manual tracings for both a normal and diseased mitral valve, and found an average difference of $0.59 \pm 0.49\text{mm}$, which is on the order of the spatial resolution of the ultrasound volumes (0.5–1.0mm/voxel).

1 Introduction

The mitral valve is a thin leaflet structure responsible for maintaining uni-directional blood flow from the left atrium to the left ventricle. Mitral valve disease, however, is one of the most prevalent among valve diseases [1], often necessitating surgical intervention. Despite this, gathering detailed information about valve dynamics on a patient-specific basis is rarely performed due to limited methods to extract valve geometry.

To better understand the mitral valve, several studies have attempted to simulate valve closure with either generic or patient-specific knowledge of the valve geometry [2,3]. Those that have used patient-specific geometry have generally done so by generating models from three-dimensional ultrasound (3DUS) [3]. The advantages of generating models from ultrasound are that ultrasound is an imaging modality that is non-ionizing and inexpensive, and is commonly used clinically to image the valve for diagnostic and interventional purposes.

The challenge with model generation from 3DUS is that a limited number of methods exist that are able to accurately and robustly segment a patient-specific 3D valve and its components with minimal input and interaction required by the user. A volumetric segmentation of the leaflets using an intensity-based level set method was shown in [4], but the segmentation method was not able to isolate the valve from the surrounding tissue and could not separate anterior from posterior leaflets upon coaptation. The work by Burlina *et al.*, 2009 uses a level set method and thin tissue detector to define the valve geometry and surroundings, but requires extensive user interaction and relies on the assumption of a planar annulus [3]. Furthermore, the segmentation is performed in only a single frame showing an open valve. The closed configuration of the valve is estimated by modeling the mechanics of the valve, but the accuracy of this approach is not validated. The work by Ionasec *et al.*, 2010 described a sophisticated system that fits a mitral and aortic model to the respective valve using machine learning techniques and a large database of manually delineated points [5]. In the context of mitral valve segmentation, however, the mitral valve model lacked fine leaflet detail due to the model being fit to only a few locations. Additionally, the model appeared to only roughly estimate the coaptation line and did not appear to be able to generate an estimate of coaptation length, which is a measure commonly used by clinicians to assess valve competency.

To address these issues, we present a 4D mitral valve segmentation method that can generate a detailed 4D patient-specific annulus and leaflet geometry. As most modeling efforts and clinical interest revolves around the closing of the valve, the method is designed to capture the valve geometry during this phase of the cardiac cycle. However, the method could easily be altered to find the leaflet geometry throughout an entire 4D ultrasound (4DUS) sequence. The method requires only the selection of two frames from an ultrasound sequence – one before the valve starts to close and another after the valve closes – and the selection of a point near the center of the closed valve. The method then automatically finds the annulus and leaflets during valve closure. By taking into account valve behavior and handling leaflet collisions, a detailed coaptation line and surface are generated. It is important to make the distinction that we are not simulating the mechanics of the valve, but rather are presenting a method to delineate the valve as seen in a 4DUS sequence. The extracted geometry is intended for inputs to mechanical models or surgical simulators, improved visualization, diagnostics, or for inverse modeling purposes.

2 Methods and Materials

2.1 Constructing Geometric Priors

The mitral valve segmentation method operates by first constructing a geometric prior of the valve which is then propagated to all other frames of interest. The valve geometry is constructed by first finding the annulus, which is then used to enforce constraints about where to find the leaflets. To find the annulus, the user-selected frame showing a closed mitral valve and the selected valve center

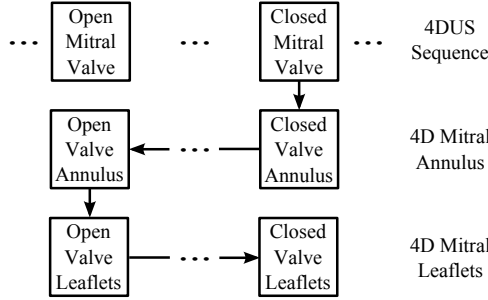


Fig. 1. Summary of 4DUS segmentation of the mitral valve. A user manually specifies the frames at the start and end of valve closure, and also a point near the center of the closed mitral valve. The annulus is found through segmentation and then tracking, starting from the closed valve and proceeding to the open. The leaflets are then found through segmentation and then tracking, starting from the open valve and proceeding to the closed.

point are used in a direct annulus segmentation method [6]. The segmented annulus is then tracked to the remaining frames of interest using a variation of the Lucas & Kanade optical flow algorithm [7], similar to that shown in [8]. The annulus in the frame selected by the user as showing the mitral valve in the open configuration before the valve begins to close is then used in a direct leaflet segmentation method to generate a detailed patient-specific triangular mesh at the leaflets [9]. The following sections describe how the mesh is evolved during valve closure such that the geometry mirrors that of the valve in the 4DUS images.

2.2 Evolving the Leaflet Mesh

The mitral valve mesh is evolved frame to frame by treating the mesh as an active surface [10], where the nodes of the mesh at the annulus are fixed according to the annulus segmentation and tracking results, and the rest of the mesh is free to evolve. The active surface is encouraged to track the top (atrial) surface of the leaflets as seen in the 3DUS images, regulated using internal edge and bending forces, and designed to handle collisions between leaflets. The active surface is then evolved according to the force equation

$$\mathbf{F}_{valve} = w_{image}\mathbf{F}_{image} + w_{edge}\mathbf{F}_{edge} + w_{bend}\mathbf{F}_{bend} + w_{collision}\mathbf{F}_{collision} \quad (1)$$

where w_{image} , w_{edge} , w_{bend} , and $w_{collision}$ are scalar weights. The nodes, \mathbf{x} , of the mesh (except for annulus nodes) are then evolved such that $\mathbf{x}^{k+1} = \mathbf{x}^k + dt\mathbf{F}_{valve}$, where dt is a time step. The annulus nodes, $\mathbf{x}_{annulus}$, are evolved according to the displacement, $\mathbf{D}_{annulus}$, and correspondence determined from tracking the annulus from frame to frame. Displacing the annulus nodes is carried out over $k_{annulus}$ time steps, such that $\mathbf{x}_{annulus}^{k+1} = \mathbf{x}_{annulus}^k + \delta_{annulus} \left(\frac{1}{k_{annulus}} \right) \mathbf{D}_{annulus}$, where $\delta_{annulus} = 1$ when $k < k_{annulus}$ and zero otherwise.

Image Force. The image force drives the mesh so that it tracks the top surface of the leaflets as seen in the 3DUS volumes. The top surface is tracked as this is the surface that defines leaflet coaptation. One component of the image force is derived from the image gradient of a Gaussian convolved thin tissue detector image [6], $\nabla G_\sigma(\mathbf{TDD})$. This component drives the valve mesh toward the center of the leaflets and helps ensure that leaflets are driven to roughly the desired location even in the presence of large displacements. We encourage the mesh to find the atrial surface of the leaflets using the gradient of the target frame’s ultrasound intensity, \mathbf{I}_{target} , along the node normal, \mathbf{N} . Having already defined the normal at each node to point toward the atrial side of the leaflets, the atrial surface of the leaflets should be at a location where $\frac{\partial \mathbf{I}_{target}}{\partial \mathbf{N}}$ is at a minimum (Figure 2). The image force is then specified as

$$\mathbf{F}_{image} = \nabla G_\sigma(\mathbf{TDD}) \cdot \mathbf{N} - \delta_{image} \frac{\partial^2 \mathbf{I}_{target}}{\partial \mathbf{N}^2} \quad (2)$$

where $\delta_{image} = 1$ when a node is at a location where $\frac{\partial \mathbf{I}_{target}}{\partial \mathbf{N}} < 0$ and zero otherwise. This is to prevent a node that is below the leaflet from being pushed even further below the leaflets.

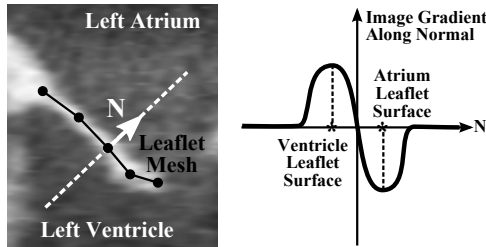


Fig. 2. Simplified depiction of the image gradient along the node normal. (Left) Ultrasound volume shown with the position of a mesh and the direction of a node normal. (Right) A simplified graph showing the gradient of ultrasound intensity along the normal. To capture the atrial side of the leaflets, nodes are forced in a direction so as to minimize the value of the gradient.

Edge Force and Artificial Strain. We maintain the geometric prior as determined from the initial leaflet segmentation by regulating the edge lengths of the mesh. We do this by imparting the edge force

$$\mathbf{F}_{edge} = \sum_{e=1}^E \delta_{edge} (L_e - L_{e,o}) \quad (3)$$

where $\delta_{edge} = \delta_c$ when $L_e - L_{e,o} < 0$ (i.e. edge is in compression) and $\delta_{edge} = \delta_t$ when $L_e - L_{e,o} > 0$ (i.e. edge is in tension). The original edge lengths are represented by $L_{e,o}$ and the number of edges attached to each node by E . For our design, we make $\delta_c < \delta_t$ as a way of penalizing longer edges over shorter.

A key issue in creating segmented models of the mitral valve is identifying the regions where the leaflets overlap when closed, as this determines the propensity for the valve to leak. This is challenging because it is impossible to discern separate leaflets in this coaptation region once the valve is closed. This necessitates tracking the leaflets as the valve closes, and approximating the stretch that ensues as the leaflets are pressurized. Studies on the mechanical properties of the valve have shown that leaflets stretch by roughly 35% in the radial direction and 20% in the circumferential direction [11] due to the pressure gradient across the valve, and because the leaflet is constrained at the annulus and by chords. As accurate chordal structure is extremely difficult to obtain from 3DUS, leaflet strain is applied artificially throughout the closure of the valve. Having determined the radial and circumferential components of each edge when the leaflet mesh is first segmented, the reference edge lengths, $L_{e,o}$ are adjusted accordingly throughout the valve. Making $\delta_c < \delta_t$ makes this artificial strain not a hard constraint but a soft constraint, and making δ_t large ensures that the leaflets do not stretch much more than the applied artificial strain. In our design, the applied strain rate is such that the leaflets are fully strained by halfway through leaflet closure to ensure that leaflets have the opportunity to coapt.

Bending Force. Due to the noise in 3DUS images, the image force has the potential to displace the leaflet mesh in ways uncharacteristic of an actual valve and in ways which cannot be corrected by the edge force. We therefore impart a bending force which tends to straighten the leaflets to further regularize the surface geometry.

$$\mathbf{F}_{bend} = \arccos(\mathbf{N}_i \cdot \mathbf{N}_j) \quad (4)$$

The bending force is computed at each edge which is shared by two triangular faces whose respective normals are \mathbf{N}_i and \mathbf{N}_j . The force is then imparted appropriately to the edge nodes and the nodes opposite the edge from each face.

Collision Force. To resolve collisions between leaflets and ultimately allow for the formation of a detailed coaptation surface, a collision force is computed between edges that are close to colliding, where potentially colliding edges are determined from a grid-based culling operation. If the distance between two edges is d_c , the force between the edges is $f_c = \exp(d_c)$. The collision force at each node is then found as

$$\mathbf{F}_{collision} = \sum_{e=1}^E f_{c,e} \quad (5)$$

where E again represents the number of edges attached to a node.

Tether Force. As previously mentioned, segmenting the chordal structure of the mitral valve from 3DUS is extremely difficult. The chords act to constrain the free edge of the leaflets. Without chords, the free edge of the leaflets has the potential to flail uncontrollably. The same is true for the leaflet mesh in the presence of the image force. We therefore found it necessary to impart a

comparatively small tethering force to the free edge of the leaflets. This is done by redirecting the image force at the free edge by adding a component to the normal that points toward the left ventricle. If the magnitude of the image force is $f_{image,o}$, which is directed along node normal \mathbf{N} , then the new image force at the free edge is

$$\mathbf{F}_{image} = \left(\frac{f_{image,o}}{2} \right) \frac{\mathbf{N} + \mathbf{A}_{valve}}{|\mathbf{N} + \mathbf{A}_{valve}|} \quad (6)$$

where \mathbf{A}_{valve} is a vector pointing along the axis of the valve toward the left ventricle which is determined when the leaflet mesh is first segmented.

3 Results

We assessed the accuracy of the presented 4D mitral valve segmentation method using retrospective ECG-gated data acquired using a transesophageal approach (iE33 Echocardiography System with a transthoracic X7-2t probe, Philips Healthcare, Andover, MA, USA). Frames at the start and end of mitral valve closure and a point near the center of the closed mitral valve were manually selected. The valve weights ($w_{image} = 0.05, w_{bend} = 0.15, w_{edge} = 0.02, w_{collision} = 2, \delta_t = 5, \delta_c = 1, dt = 1$) were tuned on a single data set to ensure stability and adequate tracking, and the method validated on two other data sets.

The results of the segmentation for the two data sets – one showing a normal mitral valve and another showing a stenotic mitral valve – at selected times during valve closure can be seen in Figure 3. To validate the segmentation method, the atrial side of the leaflets for the two data sets were manually traced by an expert in cut planes taken every 10° about the valve axis. This was done in every frame during valve closure, amounting to over 50 frames and over 1000 cut planes between the two data sets. The average differences between the manual tracings and automatically segmented meshes for the first data set (normal mitral valve), second data set (diseased mitral valve), and across both data sets were 0.84 ± 0.65 mm (RMS = 1.06mm), 0.48 ± 0.34 mm (RMS = 0.59mm), and 0.59 ± 0.49 mm (RMS = 0.77 mm), respectively. These differences were on the order of the spatial resolution of the ultrasound volumes, which were 0.5–1.0mm/voxel. This analysis makes the presented algorithm one of the most thoroughly validated 4D mitral leaflet segmentation methods.

In showing the mesh location relative to the ultrasound images in Figure 3, it can be seen that the segmentation accurately finds the valve location throughout valve closure. Additionally, the 3D mesh shows that a detailed coaptation surface is delineated. The same parameters were used for both studies, with the exception that the artificial strains for the stenotic mitral valve were made to be half those of the normal (i.e. instead of using 35% strain in the radial direction and 20% in the circumferential, we used 17% and 10%, respectively), which is an estimate for the mechanical properties of stenotic leaflet tissue.

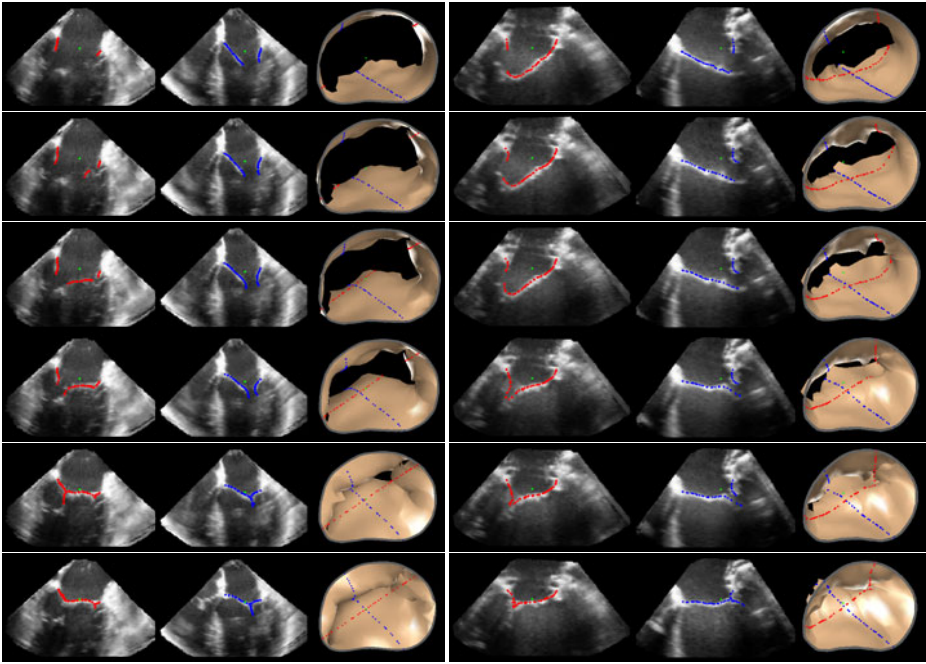


Fig. 3. Results of the mitral valve segmentation method during valve closure for a (1^{st} Column) normal and (2^{nd} Column) stenotic mitral valve. The comparison of the mesh location in the image planes (red and blue points) relative to the leaflets as seen in the images show the effectiveness of the 4D segmentation method for ultrasound.

4 Discussion

The presented mitral valve segmentation method is capable of finding the location of the mitral valve throughout valve closure, and could be modified to segment an entire cardiac cycle. The benefit of the presented method is that, with minimal user input, the method can segment the mitral valve in each frame during valve closure and is able to generate a detailed coaptation region. This is the first method capable of segmenting such patient-specific detail from a 4DUS sequence. The method benefits from assuming that a geometric prior can be very simply altered and displaced to track the desired anatomy throughout a give sequence. Were we not to assume a geometric prior and rather just segment the leaflets as seen in each separate image, the approach would be similar to that in [5], where a detailed coaptation region was not found.

As emphasized, the presented method is not performing mechanical modeling of the mitral valve. However, the method could be used in several aspects of modeling. The construction of the geometric priors could be used as a modeling input. Material properties could be determined by finding the strain values in the radial and circumferential directions that results in the best segmentation. Lastly, the 4D segmentations can be used to validate their predictions. As such, coupling segmentation and mechanical modeling could prove beneficial.

The constructed geometric priors in this study did not include the mitral valve chords. These structures are too small to be accurately segmented from three-dimensional ultrasound. For this reason, several assumptions, such as artificial strain and tethering forces, had to be made in the segmentation method to account for this lack of information. Future efforts will include specializing ultrasound data acquisition for the collection of the desired chords.

Acknowledgments. This work was supported by the U.S. National Institutes of Health under grant NIH R01 HL073647-06.

References

1. Nkomo, V., Gardin, J., Skelton, T., Gottdiener, J., Scott, C.: Enriquez-Sarano, M.: Burden of valvular heart diseases: a population-based study. *The Lancet* 368, 1005–1011 (2006)
2. Hammer, P., Vasilyev, N., Perrin, D., del Nido, P., Howe, R.: Fast image-based model of mitral valve closure for surgical planning. In: MICCAI 2008 Workshop Proceedings: Computational Biomechanics for Medicine III, New York, NY, pp. 15–26 (2008)
3. 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: Navab, N., Jannin, P. (eds.) IPCAI 2010. LNCS, vol. 6135, pp. 135–146. Springer, Heidelberg (2010)
4. Shang, Y., Yang, X., Zhu, L., Deklerck, R., Nyssen, E.: Region competition based active contour for medical object extraction. *Computerized Medical Imaging and Graphics* 32, 109–117 (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 4-D Cardiac CT and TEE. *IEEE Transactions on Medical Imaging* 29, 1636–1651 (2010)
6. Schneider, R., Perrin, D., Vasilyev, N., Marx, G., del Nido, P., Howe, R.: Mitral Annulus Segmentation from 3D Ultrasound Using Graph Cuts. *IEEE Transactions on Medical Imaging* 29, 1676–1687 (2010)
7. Lucas, B., Kanade, T.: An iterative image registration technique with an application to stereo vision. In: Proc. 7th International Joint Conference on Artificial Intelligence, Vancouver, B.C., Canada, pp. 674–679 (1981)
8. Veronesi, F., Corsi, C., Caiani, E., Sugeng, L., Weinert, L., Mor-Avi, V., Lang, R., Lamberti, C.: Semi-automatic tracking for mitral annulus dynamic analysis using real-time 3D echocardiography. *Computers in Cardiology* 33, 113–116 (2006)
9. Schneider, R., Burke, W., Marx, G., del Nido, P., Howe, R.: Modeling mitral valve leaflets from three-dimensional ultrasound. In: Metaxas, D., Axel, L. (eds.) FIMH 2011. LNCS, vol. 6666, pp. 215–222. Springer, Heidelberg (2011)
10. Terzopoulos, D., Witkin, A., Kass, M.: Constraints on deformable models: Recovering 3D shape and nonrigid motion. *Artificial Intelligence* 36, 91–123 (1988)
11. Grashow, J., Yoganathan, A., Sacks, M.: Biaxial stress–stretch behavior of the mitral valve anterior leaflet at physiologic strain rates. *Annals of Biomedical Engineering* 34, 315–325 (2006)

Vessel Connectivity Using Murray’s Hypothesis

Yifeng Jiang¹, Zhen W. Zhuang², Albert J. Sinusas^{1,2},
Lawrence H. Staib^{1,3,4}, and Xenophon Papademetris^{1,3}

¹ Department of Diagnostic Radiology

² Department of Internal Medicine Cardiology

³ Department of Biomedical Engineering

⁴ Department of Electrical Engineering,
Yale University, New Haven, CT, USA

Abstract. We describe a new method for vascular image analysis that incorporates a generic physiological principle to estimate vessel connectivity, which is a key issue in reconstructing complete vascular trees from image data. We follow Murray’s hypothesis of the *minimum work principle* to formulate the problem as an optimization problem. This principle reflects a global property of any vascular network, in contrast to various local geometric properties adopted as constraints previously. We demonstrate the effectiveness of our method using a set of microCT mouse coronary images. It is shown that the performance of our method has a statistically significant improvement over the widely adopted minimum spanning tree methods that rely on local geometric constraints.

1 Introduction

In modern vascular modeling research, a key issue is the lack of highly automated and reliable methods for *vessel connectivity*. Vascular research has been greatly facilitated by the advances of 3D imaging technologies such as MicroCT, but quantifying the images remains challenging, and constitutes a bottleneck in vascular research [1]. Although vascular image analysis has been extensively researched in the past decade [2], most previous research focuses on vessel enhancement, detection, and the segmentation and quantification of a few vessel segments or small vascular sub-trees, while vessel connectivity is seldom discussed. As a result, there are few methods capable of reconstructing complete vascular networks [3,4,5,6]. These methods are usually limited by the need for: (I.) the manual setting of numerous thresholds [3,4,5,6], (II.) the truncation of small vessels [5], (III.) manual correction [3,4]. The connectivity problem is a result of the imperfection of vessel detection. Like any feature detection results, there are missing parts and also outliers. The handling of the missing vessels (gaps) is found to be critical for connectivity. The failure of vessel detection is due to imaging artifacts/noise, missing contrast agents (*e.g.* by air bubbles), and morphology beyond detection models.

Previous work in vessel connectivity [2] largely depends on local image or vascular context. Many methods (*e.g.* [4,6]) connect vessels based on a set of intuitive rules using local vessel orientation and distances. A set of thresholds are

often employed to decide connectivity. Minimum spanning tree (MST) methods offer a more principled approach. Various distance metrics (thus cost functions), *e.g.* Euclidean distance [7], Mahalanobis distance [8], and a mixture of Euclidean distance and vessel orientation difference [3] has been developed for this approach. Although their cost functions are global, they have optimal substructure and still establish the connections based on local context. Some methods use “gap filling” techniques which more systematically explore the local image context. They adjust the distance between nodes [3,8], or estimate missing nodes by tensor voting and adaptive thresholding, *etc.* However, the use of local context fails as the gaps get larger. Furthermore, intuitively defined thresholds and distance metrics also have difficulty coping with spatially varying gap geometries.

Our method: In this paper, we leverage physiologically-based global properties of vasculature trees to guide the estimation of vessel connectivity. There has been little work in vascular image analysis that attempts to explore the wealth of knowledge accumulated in physiology research, such as (I.) the morphometric statistics of vasculature experimentally measured; (II.) measured fractal properties of the vascular trees; (III.) the minimum work principle [9]; (IV.) The scaling laws among perfusion, mass, and vascular morphology; (V.) The flow and pressure within the vascular system. In this paper, we start with *the minimum work principle* (MWP), Murray’s hypothesis, as introduced in Section 2.2, which is conceptually simple and also comparatively easy for mathematical formulation. A few previous methods on vessel image segmentation were also inspired by this principle [10,11], but the resulting formulations were ad-hoc and the results were not quantitatively evaluated. We present a formal derivation by strictly following this principle and quantitative validation on its utility for determining vessel connectivity.

2 Method

2.1 The Vessel Connectivity Problem

We consider the problem as: **given** vessel centerline points \mathcal{V} , and the vessel radii \mathcal{R} at corresponding locations of \mathcal{V} , **find** the connections \mathcal{E} among \mathcal{V} so that the vascular tree $\mathcal{T} = (\mathcal{V}, \mathcal{E}, \mathcal{R})$ is reconstructed. In many cases, the vascular tree root v^r is also given. In this paper, we assume that the vascular network under consideration has a tree structure. We intend to design a cost function $\mathbf{F}(\cdot)$ which measures the truthfulness of the morphology and topology of a vascular tree. The connectivity problem can then be written as

$$\hat{\mathcal{E}} = \arg \min_{\mathcal{E}} \mathbf{F}(\mathcal{E}, \mathcal{V}, \mathcal{R}, v^r). \quad (1)$$

In an image analysis scenario, the vessel points actually detected, \mathcal{V}^* (with corresponding radii \mathcal{R}^*), always contain a set of outlier points $\mathcal{V}^o = \mathcal{V}^* \setminus \mathcal{V}$ (false positive, with radii \mathcal{R}^o) and at the same time miss a set of points $\mathcal{V}^m = \mathcal{V} \setminus \mathcal{V}^*$ (false negative, with radii \mathcal{R}^m), due to various reasons including image noise,

artifacts, unrealistic modeling assumed in detection, *etc.* The vessel radius estimation will not be accurate either, but for simplicity, this is not discussed in this paper. Therefore the real vessel connectivity problem is a Steiner tree problem

$$(\hat{\mathcal{E}}, \hat{\mathcal{V}}^m, \hat{\mathcal{V}}^o) = \arg \min_{(\mathcal{E}, \mathcal{V}^m, \mathcal{V}^o)} \mathbf{F}(\mathcal{E}, \mathcal{V}^* \setminus \mathcal{V}^o \cup \mathcal{V}^m, \mathcal{R}^* \setminus \mathcal{R}^o \cup \mathcal{R}^m, v^r). \quad (2)$$

The topology of vasculature is often more important than the accurate locations of connection. In this paper, we focus on recovering the vascular topology without estimating \mathcal{V}^m , and also leave the estimation of \mathcal{V}^o to future work.

2.2 Murray’s Hypothesis

Nature pursues optimality in all her workings. Murray [9] was among the first to quantify this hypothesis for vascular systems. He proposed the Minimum Work Principle (MWP): the sum of the work overcoming flow resistance for perfusion and the metabolic consumption of the blood is naturally minimized. He defined the *work* [9] as

$$\mathbf{P} = \mathbf{P}^f + \mathbf{P}^b = p \cdot f + b \cdot \text{vol} \quad (3)$$

$$= (f^2 \cdot l \cdot 8\eta) / (\pi r^4) + b \cdot l \pi r^2, \quad (4)$$

where \mathbf{P}^f denotes power dissipation caused by flow friction, and \mathbf{P}^b represents the consumption of metabolism. Eqn. (3) applies to the whole vasculature, where p and f are pressure and flow at the entrance of vasculature, *vol.* is the volume of vessel, and b is the metabolic cost of blood in *ergs/sec./cm³*. Eqn. (4) calculates the work for one specific vessel segment, following Poisseuille’s equation, where r is the vessel diameter, η is blood viscosity, and l is the length of the vessel segment. Using (4), it is straightforward to derive the optimal relationship between f and r for a vessel segment with fixed length l :

$$\partial \mathbf{P} / \partial r = 0 \Rightarrow f = r^3 \frac{\pi}{4} \sqrt{b / \eta}. \quad (5)$$

The above model has been justified in a number of ways [9]. The relation between branching angle and radii derived from this hypothesis, has been found to closely conform to the morphologies of various biological vascular systems including plants. Although the model does not conform to anastomoses in vasculature regions such as the Circle of Willis in cerebral vasculature, it is applicable to most of the vascular system.

2.3 Problem Formulation

If we assume Murray’s \mathbf{P} (3,4) measures the quality of an estimated vasculature, *i.e.* it can serve as the $\mathbf{F}(\cdot)$ in Eqn. (1), our vessel connectivity problem becomes

$$\mathcal{E} = \arg \min_{\mathcal{E}} \sum_{k=1}^N \underbrace{[(f^2(e_k, \mathcal{E}) \cdot l(e_k) \cdot 8\eta) / (\pi r^4(e_k))]}_{\mathbf{P}_k^f} + \underbrace{[b \cdot l(e_k) \pi r^2(e_k)]}_{\mathbf{P}_k^b}, \quad (6)$$

where $\mathbf{P}_k = \mathbf{P}_k^f + \mathbf{P}_k^b$ is the total work consumption for one vessel segment (Eqn.4), $e_k \in \mathcal{E}$, $e_k = (v_i, v_j)$, $v_i, v_j \in \mathcal{V}$, represents a vessel segment, starting at v_i and terminating at v_j , with $N = |\mathcal{E}|$ as the total number of segments, $f(\cdot)$ measures the flow through a vessel segment, $l(\cdot)$ calculates the segment length, and $r(\cdot)$ denotes the segment radius, which corresponds to the radius $r \in \mathcal{R}$ associated with the second point $v_j \in e_k$ or the mean of both points.

Formulation (MST+) assuming optimal $f \sim r$: If the optimal relation between f and r (5) is assumed, it’s straightforward to simplify (6) and yield an $r^2(e_k)$ weighted MST (we label it as **MST+**) formulation as:

$$\mathcal{E} = \arg \min_{\mathcal{E}} \sum_{k=1}^N l(e_k)r(e_k)^2. \tag{7}$$

This indicates that published MST variants [3,7,8] can also be derived from MWP with certain simplifying assumptions (*e.g.* $r = 1$), which was not previously demonstrated.

Formulation with flow conservation constraint: Eqn. (7) does not impose the flow conservation between parent and their daughter branches. In the original form (6) this constraint is embedded in the calculation of $f(e_k, \mathcal{E})$ which is unfortunately computationally intractable. In this paper, we impose this constraint by converting the resistive tree into a parallel network. As illustrated in Fig. 1, resistor R_0 in a tree network is split into R_0^1 and R_0^2 , which are determined by their flow, I_1 and I_2 respectively:

$$R_0^1 = I_0/I_1 \cdot R_0, \quad R_0^2 = I_0/I_2 \cdot R_0, \quad \text{where } I_0 = I_1 + I_2. \tag{8}$$

We apply this simple conversion repetitively starting from the branch ends of the vascular tree, $v_m^e, m = 1, 2 \dots M$, all the way to its root, v^r , which makes the vascular structure in between fully parallel. Then the total work dissipated on flow friction is simply the sum of the work through each path,

$$\mathbf{P}^f = \sum_{m=1}^M f_m^2 \sum_{e_k \in \mathcal{P}(v_m^e, v^r, \mathcal{E})} \frac{f(e_k, \mathcal{E})}{f_m} \cdot \frac{l(e_k) \cdot 8\eta}{\pi r^4(e_k)}, \tag{9}$$

where f_m is the flow at branch end v_m^e , and $\mathcal{P}(v_m^e, v^r, \mathcal{E})$ denotes the flow path from v_m^e to vascular root v^r . This formulation explicitly involves the connectivity of the vasculature by introducing $\mathcal{P}(v_m^e, v^r, \mathcal{E})$. Assuming the optimal $f \sim r$ (5), the cost function using (9) is simplified to be

$$\mathcal{E} = \arg \min_{\mathcal{E}} \underbrace{\sum_{m=1}^M \frac{r(e_m)^3}{2}}_{SPT} \sum_{e_k \in \mathcal{P}(v_m^e, v^r, \mathcal{E})} \frac{l(e_k)}{r(e_k)} + \underbrace{\sum_{k=1}^N l(e_k)r(e_k)^2}_{MST}, \tag{10}$$

which is a mixture of the shortest path like tree (SPT) and MST.

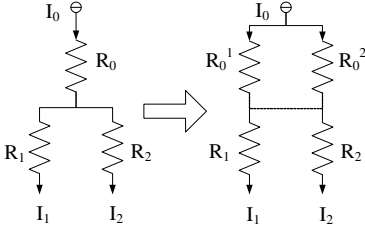


Fig. 1. Converting a tree-structure resistive circuit to a parallel one

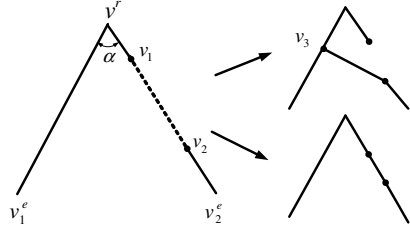


Fig. 2. Left: The simplest tree with a gap (v_1, v_2) . **Right:** two possible solutions.

It is not difficult to show that the formulation with flow constraint (10) has advantages over the former MST formulation (7) on bridging the gaps between detected vessel segments. The simplest tree in Fig. 2 is used for illustration, with a root node v^r and two branches ending at v_1^e and v_2^e . The detected vessel points along both branches are assumed to be continuous except for the gap between v_1 and v_2 , where points are totally missing. Consider the effect of this gap on connecting v_2 to v_1 , as desired, or to some point v_3 along branch (v^r, v_1^e) by mistake. We assume that the radii $r(v_1^e) = r(v_2^e) = 1$ for simplicity which results in (7) becoming a pure MST formulation. In this case, the optimal angle between the branches $\alpha \approx 75^\circ$ and often smaller in real data [9]. The MST cost function (7) becomes $\sum_{k=1}^N l(e_k)$ and the proposed formulation (Eqn. 10) equals $\sum_{m=1}^M \frac{1}{2} \sum_{e_k \in \mathcal{P}(v_m^e, v^r, \mathcal{E})} l(e_k) + \sum_{k=1}^N l(e_k)$. In order to establish the right connection, connecting (v_1, v_2) must have a lower cost than connecting (v_3, v_2) . Using MST (7), we need

$$l(v_1, v_2) < \min_{v_3} l(v_3, v_2), \tag{11}$$

whereas using formulation (10), it is not difficult to derive the constraint:

$$l(v_1, v_2) < \min_{v_3} [l(v^r, v_3)/3 + l(v_3, v_2)]. \tag{12}$$

This means the upper bound of the allowed gap $l(v_1, v_2)$ is *always* greater when using the second cost function.

2.4 Optimization

The optimization for the proposed formulation (10), despite its simple form, turns out to be non-trivial. In this paper, we approximate it by assuming that radius r_m is given and remains the same for all branch ends $\{v_m^e\}$, which is the size of an image voxel r_{vox} . Note that not all points of radius r_{vox} belong to $\{v_m^e\}$. $\{v_m^e\}$ is still a result of the optimization. Then the first term in (10) is further approximated by a pure cost of shortest path tree (*i.e.* making every node a branching end) multiplied by an coefficient $\lambda \in (\frac{M}{|V|}, 1)$. The cost is then approximated by

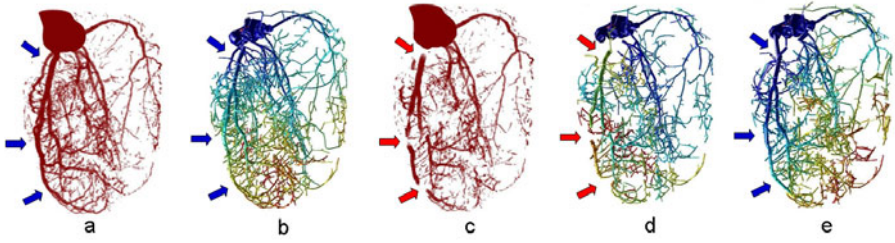


Fig. 3. A visual comparison of reconstructed trees: (a) The original image; (b) Manually validated vascular tree; (c) The image contaminated by gaps; (d) The tree constructed by MST; and (e) The tree constructed using our cost function (10). The warmer colors indicate deeper branching levels of vessel segments.

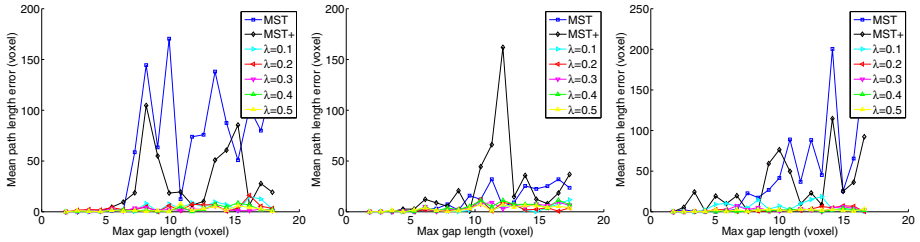


Table 1. Significance

Gap	$P_{\lambda=0.1}$	$P_{\lambda=0.3}$	$P_{\lambda=0.5}$
6%	0.154	0.100	0.102
8%	0.064	0.061	0.062
10%	0.029*	0.028*	0.029*
12%	0.003*	0.003*	0.002*
14%	0.004*	0.003*	0.003*
16%	0.026*	0.021*	0.023*
18%	0.000*	0.000*	0.000*
20%	0.065*	0.031*	0.031*

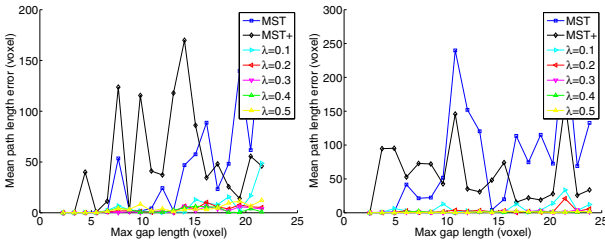


Fig. 4. A quantitative comparison of reconstructed trees (Section 3). Each subfigure presents the error obtained from one vascular tree. **MST**: minimizing MST cost function; **MST+**: minimizing (7); The others: minimizing the proposed function (10) (approximated by (13) with different λ). P values from t tests comparing the proposed function (10) with MST are in Tab. 1. * marks significant results.

$$\mathbf{P}^* = \lambda \cdot \frac{r_{vox}^3}{2} \sum_{j=1}^N \sum_{e_i \in \mathcal{P}(v_j, v^r, \mathcal{E})} \frac{l(e_i)}{r(e_i)} + \sum_{k=1}^N l(e_k) r^2(e_k). \tag{13}$$

Note that as a tree, $N = |\mathcal{E}|$ is also the number of all vertices besides v^r . This expression is equivalent to the so-called cable-trench problem [12]. Although it

is still a difficult problem (NP-complete), we found that it can be effectively approximated by Prim’s algorithm with distance between vertices defined in (13). As a greedy algorithm, when implemented in Matlab it can establish the connection for 10,000 vessel points within 10 seconds on a state-of-the-art PC. Although (10) can be better approximated by more sophisticated algorithms, this formulation is shown to be adequate in the results described below.

3 Experimental Results

We evaluate the algorithm performance by examining how well the connectivity of a set of real mouse coronary arterial trees can be recovered. These trees were manually validated and gaps at known locations with known sizes were inserted.

Vascular trees: Five healthy C57Bl/6 mouse coronary arterial trees were extracted from their ex-vivo microCT images with manual image processing by an expert. The image size is $512 \times 512 \times 742$ with voxel resolution of $16\mu m$. The data were first manually edited within the Volume Viewer of the GE Advantage Workstation to remove venous and other outlier structures. They were then down-sampled (to resolution of $32\mu m$) to reduce the computational demand of validation. The resulting images were skeletonized by image thinning with manual adjustment of the threshold.

Evaluation: We created gaps by masking out part of \mathcal{V} using 500 spherical balls from any original phantom tree $\mathcal{T} = (\mathcal{V}, \mathcal{E}, \mathcal{R}, v^r)$. The simulated gaps are similar to those common ones in real images (*e.g.* caused by air bubbles or weak vessel detection). The ball diameters ranged from 2% to 20% of the width of the bounding box of the whole vasculature. The resulting gaps had maximum lengths around 0.06–0.6 mm in physical space, or 2–20 voxel widths in the image space. For each ball size, 100 random configurations of the ball locations are used for evaluation. The ground truth connectivity \mathcal{E}^- for the remaining vessel points and associated radii $(\mathcal{V}^-, \mathcal{R}^-, v^r)$ can be obtained from \mathcal{E} . $(\mathcal{V}^-, \mathcal{R}^-, v^r)$ were input into different algorithms to find $\hat{\mathcal{E}}^-$. Directly comparing \mathcal{E}^- and $\hat{\mathcal{E}}^-$ to counting the number of wrong connections is inadequate to reflect the significance of error in terms of the function of a vascular tree. We consider the path lengths from tree root v^r to each vessel point as better error measures, because they are associated with the tree function, and will also be altered by erroneous connections. We define the error as $\mathbf{E} = \frac{1}{|\mathcal{V}^-|} \sum_{v_i \in \mathcal{V}^-} |l(\mathcal{P}(v_i, v^r, \hat{\mathcal{E}}^-)) - l(\mathcal{P}(v_i, v^r, \mathcal{E}^-))|$, where $\mathcal{P}(\cdot)$ finds the shortest path between two nodes within a tree, and $l(\cdot)$ measures its length. For each ball size, \mathbf{E} is further averaged over random configurations of ball locations.

Results: The experimental results are plotted in Fig. 4. The error \mathbf{E} generally increases when the maximum gap length increases. It is observed that when the maximum gap length is greater than 8% of the vasculature width, the errors resulting from minimizing our proposed cost function (10) are significantly lower

than those from minimizing the pure MST cost function, with λ in a comparatively wide range, (0.1, 0.5). It is also observed that MST+ formulation (7) leads to a similar level of error to the pure MST. This is not surprising as discussed earlier in Section 2.3. In Fig. 3 we also plot a sample result of both MST and our algorithm for a visual comparison. To create large gaps but keep the majority of the vasculature, only 100 balls of random diameters from 2 to 20 voxel width were used. It is obvious that at the places where significant gaps exist, our cost function recovers far more better connectivity than MST.

4 Conclusions and Future Work

We hypothesized that global vascular properties would provide better guidance for vessel connectivity than those local ones. In this paper, we follow Murray's hypothesis of the MWP (9) to formulate the vessel connectivity as an optimization problem. A cost function (10) that naturally imposes flow conservation is derived, and a greedy algorithm is also proposed for optimization. We evaluate the performance of our algorithm quantitatively using real vascular image data and demonstrate that it can achieve significant improvements in comparison with the results of MST, whose variants were widely adopted (3,7,8).

Others have attempted to apply similar principles. Bruyninckx *et al.* (10) employed Murray's law for extracting vascular trees from images. As briefly mentioned in Section 2.2, Murray's law is derived from the MWP, describing a local vascular property. For the global constraints they minimized the total vascular volume, which is half of MWP (the second term in (3)). Previously we also tried to employ the MWP to connect vessels (11). The major difference is that a pseudo vascular resistance was used to formulate the cost function, instead of deriving it directly from the MWP. The algorithms developed in (10,11) are far more computationally demanding than the greedy one proposed in this paper and there were no quantitative evaluations conducted for vessel connectivity.

We conceptually validated the idea of using Murray's hypothesis for vessel connectivity. Future work will include refining the optimization process and developing other formulations of the vessel connectivity problem using MWP, including using perfusion volume constraint for outlier detection (estimation of \mathcal{V}^o), and the extraction of multiple vascular trees (e.g. arterial and venous trees).

Acknowledgements. This work was supported by the NIH under grants R01 HL065662 (AJS), R01 EB006494 (XP), and CTSA Grant UL1 RR024139.

References

1. Zagorchev, L., Oses, P., Zhuang, Z.W., Moodie, K., Mulligan-Kehoe, M., Simons, M., Couffinal, T.: Micro computed tomography for vascular exploration. *Journal of Angiogenesis Research* 2(1), 7 (2010)
2. Lesage, D., Angelini, E., Bloch, I., Funke-Lea, G.: A review of 3D vessel lumen segmentation techniques: Models, features and extraction schemes. *Med. Image Anal.* 13(6), 819–845 (2009)

3. Szymczak, A., Stillman, A., Tannenbaum, A., Mischaikow, K.: Coronary vessel trees from 3d imagery: a topological approach. *Med. Image Anal.* 10(4), 548–559 (2006)
4. Lee, J., Beighley, P., Ritman, E., Smith, N.: Automatic segmentation of 3D micro-CT coronary vascular images. *Med. Image Anal.* 11(6), 630–647 (2007)
5. Wischgoll, T., Choy, J., Ritman, E., Kassab, G.: Validation of image-based method for extraction of coronary morphometry. *Ann. Biomed. Eng.* 36(3), 356–368 (2008)
6. Bauer, C., Pock, T., Sorantin, E., Bischof, H., Beichel, R.: Segmentation of interwoven 3d tubular tree structures utilizing shape priors and graph cuts. *Med. Image Anal.* 14(2), 172–184 (2010)
7. Bullitt, E., Aylward, S., Liu, A., Stone, J., Mukherji, S., Coffey, C., Gerig, G., Pizer, S.: 3D graph description of the intracerebral vasculature from segmented MRA and tests of accuracy by comparison with x-ray angiograms. In: Kuba, A., Sámal, M., Todd-Pokropek, A. (eds.) *IPMI 1999*. LNCS, vol. 1613, pp. 308–321. Springer, Heidelberg (1999)
8. Jomier, J., LeDigarcher, V., Aylward, S.: Automatic vascular tree formation using the mahalanobis distance. In: Duncan, J., Gerig, G. (eds.) *MICCAI 2005*. LNCS, vol. 3750, pp. 806–812. Springer, Heidelberg (2005)
9. Murray, C.: The physiological principle of minimum work. I. The vascular system and the cost of blood volume. *Proc. Natl. Acad. Sci.* 12(3), 207–214 (1926)
10. Bruyninckx, P., Loeckx, D., Vandermeulen, D., Suetens, P.: Segmentation of liver portal veins by global optimization. In: *Proc. of SPIE*, vol. 7624, p. 76241Z (2010)
11. Jiang, Y., Zhuang, Z., Sinusas, A., Papademetris, X.: Vascular tree reconstruction by minimizing a physiological functional cost. In: *CVPRW*, pp. 178–185. IEEE (2010)
12. Vasko, F., Barbieri, R., Rieksts, B., Reitmeyer, K., Stott, K., et al.: The cable trench problem: combining the shortest path and minimum spanning tree problems. *Comput. Oper. Res.* 29(5), 441–458 (2002)

Segmentation of Nerve Bundles and Ganglia in Spine MRI Using Particle Filters

Adrian Dalca¹, Giovanna Danagoulian², Ron Kikinis^{2,3},
Ehud Schmidt², and Polina Golland¹

¹ MIT Computer Science and Artificial Intelligence, Cambridge, MA, USA

² Department of Radiology, Brigham and Women's Hospital, Boston, MA, USA

³ Surgical Planning Laboratory, Brigham and Women's Hospital, Boston, MA, USA

Abstract. Automatic segmentation of spinal nerve bundles that originate within the dural sac and exit the spinal canal is important for diagnosis and surgical planning. The variability in intensity, contrast, shape and direction of nerves seen in high resolution myelographic MR images makes segmentation a challenging task. In this paper, we present an automatic tracking method for nerve segmentation based on particle filters. We develop a novel approach to particle representation and dynamics, based on Bézier splines. Moreover, we introduce a robust image likelihood model that enables delineation of nerve bundles and ganglia from the surrounding anatomical structures. We demonstrate accurate and fast nerve tracking and compare it to expert manual segmentation.

Keywords: nerve bundles, tracking, segmentation, particle filter.

1 Introduction

Mapping and localization of nerve pathways is essential for diagnosis of spinal pathologies, treatment planning, and image-guided interventions. Recent developments in high-resolution MRI have enabled visualization of the nerve bundles within the dura, as they pass through the foramen, and exit the vertebral canal [16]. The bundles exhibit good contrast with fluids and bone, but are often of similar intensity to that of marrow and muscle. Manual segmentation of nerves and ganglia is quite challenging and time-consuming. In this paper, we develop and demonstrate a method for automatic segmentation of these nerves and ganglia in high-resolution MRI that requires minimal input from an expert.

Nerve bundles and ganglia can be observed in high-resolution (0.3-0.6 mm voxels) MRI, as illustrated in Fig. 1. They start inside the spinal canal as dark grey bundles. The intensity contrast changes along the nerve; the neighboring tissues include cerebral spinal fluid, bone and other nerves. The bundles approximate cylinders, but change shape in the presence of a pathology such as disc herniation. As the bundles turn and exit through the foramen, they grow thicker, less regular and have lower contrast with their surroundings (e.g., fat, muscle or bone) as the nerve ganglia form. Automated nerve segmentation promises to significantly improve image-based diagnosis and therapy by providing fast

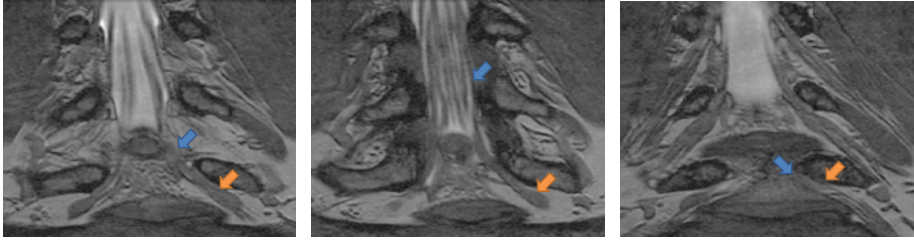


Fig. 1. Three slices from an example herniated disk MRI. All arrows point to the same nerve bundle. Blue arrows show examples of poor contrast between the nerve and surrounding tissue; orange arrows indicate the thickening of the nerve into a ganglion.

tracing of the nerve bundles. In fact, the amount of time and effort currently required for manual tracing of the nerves precludes the practical construction of nerve maps, which could benefit surgery and minimally invasive interventions by reducing procedure duration and complications and improving outcomes of treatment of several degenerative conditions and spinal trauma.

Prior work in segmentation of tubular structures, such as blood vessels, is clearly relevant to this task. Region-growing approaches, such as classical region-growing [5], competitive region-growing [20] and wave propagation [6], have been used successfully for vessel segmentation. Unfortunately, these methods suffer from leakage, where segmentation “leaks” into nearby structures, in areas with lack of contrast [11]. Active contours that evolve an initial boundary to segment the tubular targets offer a model-based approach [13,14]. Variational and level-set formulations have been successfully used for vascular segmentation [7,13]. However, these methods need good initializations and suffer from many local minima [11]. Moreover, false positive rate (or severe leakage) is certain to limit the usefulness of such methods in our application.

An alternative approach is to first extract the centerline, and to fill out the segmentation as a subsequent step. The target is modeled as a tubular structure – most often via circular cross-sections of inscribed cylinders or spheres – with a centerline in 3D [19]. In practice, these methods require a fair amount of user interaction to often re-seed tubes or branches [11]. Correction and re-centering during centerline propagation are affected substantially by neighboring tissues of similar intensities. Some methods require two end-points and employ variants of minimal path extraction [4,10], which often suffer from shortcut paths [12]. The high degree of required user input and the high false-positive rate in low contrast structures make these approaches inappropriate for our specific task. A recent study has demonstrated the feasibility of using DTI and fiber tracking of lumbar nerves [2] to estimate differences of diffusion parameters between healthy and herniated discs. Our method can be readily extended to include diffusion information by incorporating directional information in the likelihood function.

In this paper, we present a tracking approach based on particle filtering, also known as sequential Monte-Carlo tracking. Tracking has also been demonstrated previously for segmentation of tubular structures [8,11,15]. Most vessel tracking

methods model the state as a cross-sectional ellipse [3,8] or as a cylindroid [19]. In tracking nerve bundles, the regions of low contrast require the state to capture substantially longer segments of the track than what is represented by a cross-section. In addition, nerves tend to change direction, often sharply, which necessitates a use of more complex descriptors than cylinders. Enabling multiple hypotheses, flexible dynamics and diverse likelihood models, particle filtering is an ideal approach for our task.

To address the challenges of nerve tracking, we define a rich particle representation that captures the geometric behaviour of the nerve bundles. We use a Bézier spline [18] centerline with a quadratic radius function to characterize a nerve bundle. We devise a dynamics model for particle updates that enforces continuity and smoothness. Furthermore, we define an image likelihood model that compares gradient fields and intensities of predicted patches with image observations to evaluate the posterior distribution of the particles' importance. Once tracking is completed, we remove spurious segmentations by measuring the quality of the entire tract. We demonstrate successful segmentations of nerves and evaluate them relative to expert manual segmentations. To the best of our knowledge, this is the first automatic segmentation of nerve bundles and ganglia.

2 Methods

Particle Filters. We start by providing a brief review of particle filters [17]. We let \mathbf{h}_t be the particle, or state, representation at step t of the tracking algorithm, and assume the states respect a first order Markov chain, i.e.,

$$p(\mathbf{h}_t | \mathbf{h}_{1:t-1}) = p(\mathbf{h}_t | \mathbf{h}_{t-1}), \quad (1)$$

where $\mathbf{h}_{1:t-1}$ denotes the state history for step t . We let \mathbf{z}_t be the image-based observation at step t , and further assume that given the state, the observations at different time points are independent:

$$p(\mathbf{z}_{1:t} | \mathbf{h}_{1:t}) = p(\mathbf{z}_t | \mathbf{h}_t) \cdot p(\mathbf{z}_{1:t-1} | \mathbf{h}_{1:t-1}). \quad (2)$$

A step of the tracking algorithm estimates the posterior distribution $p(\mathbf{h}_t | \mathbf{z}_{1:t})$, represented non-parametrically via a set of K weighed samples $\{\mathbf{h}_t^{(k)}, w_t^{(k)}\}_{k=1}^K$. It can be readily shown that

$$p(\mathbf{h}_t | \mathbf{z}_{1:t}) \propto p(\mathbf{z}_t | \mathbf{h}_t) \cdot p(\mathbf{h}_t | \mathbf{z}_{1:t-1}), \quad (3)$$

i.e., the particle distribution depends on a likelihood function $p(\mathbf{z}_t | \mathbf{h}_t)$ and a prior term $p(\mathbf{h}_t | \mathbf{z}_{1:t-1})$ [17].

A particle filter maintains the posterior distribution (3) by generating a set of samples $\{\mathbf{h}_t^{(k)}, w_t^{(k)}\}_{k=1}^K$ at step t from the set $\{\mathbf{h}_{t-1}^{(k)}, w_{t-1}^{(k)}\}_{k=1}^K$ generated at step $t-1$. Specifically, a sample $\langle \mathbf{h}_{t-1}, w_{t-1} \rangle$ is drawn from the set $\{\mathbf{h}_{t-1}^{(k)}, w_{t-1}^{(k)}\}$ according to the weights $\{w_{t-1}^{(k)}\}$, and is propagated to become $\langle \mathbf{h}_t^{(k)}, w_t^{(k)} \rangle$. The state vector $\mathbf{h}_t^{(k)}$ is sampled from $p(\mathbf{h}_t | \mathbf{h}_{t-1})$, and its weight $w_t^{(k)}$ is computed

by rescaling w_{t-1} with likelihood $p(\mathbf{z}_t|\mathbf{h}_t^{(k)})$. The weights are normalized at each step to sum to 1. It can be shown that the resulting sample set $\{\mathbf{h}_t^{(k)}, w_t^{(k)}\}_{k=1}^K$ is an accurate representation of the state distribution $p(\mathbf{h}_t|\mathbf{z}_{1:t})$.

In the remainder of this section, we define our particle model \mathbf{h}_t for the nerve bundle, our dynamics model $p(\mathbf{h}_t|\mathbf{h}_{t-1})$, and our likelihood measure $p(\mathbf{z}_t|\mathbf{h}_t)$. Together, these elements fully define the tracking algorithm.

Particle Representation for Nerve Tracks.

We model each particle as a tubular structure around a centerline curve in 3D. We design the centerline as a Bézier curve and form the tubular structure by introducing a radius function, as illustrated in Fig. 2. A n^{th} degree Bézier curve [18] is defined by $n + 1$ control points. The first and last control points define the end-points of the curve. The interior control points can be thought of as “pulling” the curve towards them. We choose to work with cubic curves, i.e.,

$$\mathbf{c}(\tau) = (1 - \tau)^3\mathbf{p}_0 + 3(1 - \tau)^2\tau\mathbf{p}_1 + 3(1 - \tau)\tau^2\mathbf{p}_2 + \tau^3\mathbf{p}_3, \tag{4}$$

where $\tau \in [0, 1]$ is the parameterization variable.

We allow the radius function $r(\cdot)$ to vary quadratically along the segment, and also define it via a Bézier curve:

$$r(\tau) = (1 - \tau)^2r_0 + 2(1 - \tau)\tau r_1 + \tau^2r_2, \tag{5}$$

for some control points r_0, r_1 , and r_2 . In addition, we maintain the mean image intensity μ inside the segment. The state vector

$$\mathbf{h} = (\mathbf{p}_0, \mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3, r_0, r_1, r_2, \mu) \tag{6}$$

fully describes the corresponding segment. This construction can handle tubular structures with variable directionality, thickness and contrast, such as nerve bundles and ganglia.

Dynamics model. We now describe the construction of the state vector \mathbf{h}_t from the state vector \mathbf{h}_{t-1} generated in a previous step of the algorithm. This step corresponds to sampling the probability $p(\mathbf{h}_t|\mathbf{h}_{t-1})$.

We set the first centerline control point of \mathbf{h}_t to the last control point of \mathbf{h}_{t-1} : $\mathbf{p}_{0,t} = \mathbf{p}_{3,t-1}$, which ensures continuity of the track. We place $\mathbf{p}_{1,t}$ along the line $(\mathbf{p}_{2,t-1}\mathbf{p}_{3,t-1})$. This maintains smoothness during the transition from the previous to the current particle, since for Bézier curves, the tangent of a curve at an endpoint \mathbf{p}_0 is along the vector $(\mathbf{p}_0, \mathbf{p}_1)$, and similarly the tangent at point \mathbf{p}_3 is along the vector $(\mathbf{p}_2, \mathbf{p}_3)$. The distance ℓ between $\mathbf{p}_{0,t}$ and $\mathbf{p}_{1,t}$ is drawn uniformly from $[0, L]$ where L is a parameter of the algorithm. Formally,

$$\mathbf{p}_{1,t} = \mathbf{p}_{0,t} + \ell\hat{\mathbf{n}},$$

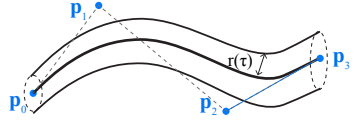


Fig. 2. Nerve segment (particle) is defined as a Bézier curve centerline with four control points and a radius function $r(\cdot)$

where $\hat{\mathbf{n}}$ is the unit vector in the direction of $(\mathbf{p}_{2,t-1}, \mathbf{p}_{3,t-1})$.

We draw the direction of $(\mathbf{p}_{1,t}, \mathbf{p}_{2,t})$ from a von Mises-Fisher distribution on the unit sphere, centered on the direction of the initial tangent $(\mathbf{p}_{0,t}, \mathbf{p}_{1,t})$ with concentration $1/\sigma_\alpha^2$ (we call this new direction $\hat{\mathbf{n}}_{01} + \Delta\hat{\mathbf{n}}_{12}$). We then sample the distance ℓ of $\mathbf{p}_{2,t}$ from $\mathbf{p}_{1,t}$ uniformly from $[0, L]$. Formally,

$$\mathbf{p}_{2,t} = \mathbf{p}_{1,t} + \ell_2(\hat{\mathbf{n}}_{01} + \Delta\hat{\mathbf{n}}_{12}).$$

Using the same procedure we generate $\mathbf{p}_{3,t}$ based on $\mathbf{p}_{2,t}$ and vector $(\mathbf{p}_{1,t}, \mathbf{p}_{2,t})$.

We set $r_{0,t} = r_{2,t-1}$ to maintain continuity of the radius function. We sample a distance $d \in (0, 1)$, and we set $r_{1,t}$ to the y-coordinate of a control point which is d away from $r_{0,t}$ along the line $(r_{1,t-1}, r_{2,t-1})$. We choose $r_{2,t}$ from a Gaussian distribution with mean $r_{1,t}$ and variance σ_r^2 . The intensity parameter μ_t is propagated via a Gaussian distribution with variance σ_μ^2 .

This construction depends on four parameters: L controls the length of the particle, σ_α^2 determines the variation in the particle curvature, σ_r^2 describes the range of the radius, and σ_μ^2 captures the intensity variation.

Likelihood Measure. Now we focus on the likelihood $p(\mathbf{z}_t|\mathbf{h}_t)$. The contrast between the nerve bundle and surrounding tissue may change along the track, rendering ribbon measures [8][15] inappropriate. However, it still creates image gradients normal to the centerline. Therefore, to measure alignment of an observed image patch I with a hypothesis \mathbf{h}_t , we simulate an image patch with a white nerve segment described by the particle \mathbf{h}_t on a black background. We compare the directions of its gradients \mathbf{g}^h with those of the observed patch \mathbf{g}^I . We avoid using the gradient magnitude, since false hypotheses with partial but very strong contrasts are abundant. We express the distance d_{∇}^2 between the normalized particle gradient $\hat{\mathbf{g}}^h$ and the normalized observed gradient $\hat{\mathbf{g}}^I$ via their cross product:

$$d_{\nabla}^2(\mathbf{h}_t, I) = \frac{1}{|V(\mathbf{h}_t)|} \sum_{v \in V(\mathbf{h}_t)} (\hat{\mathbf{g}}_v^h \times \hat{\mathbf{g}}_v^I)^2 = \frac{1}{|V(\mathbf{h}_t)|} \sum_{v \in V(\mathbf{h}_t)} \sin^2(\hat{\mathbf{g}}_v^h, \hat{\mathbf{g}}_v^I), \quad (7)$$

where $V(\mathbf{h}_t)$ is the set of voxels in the simulated patch that belong to the predicted nerve segment and v is an image voxel.

To capture relatively constant nerve intensity, we include a term that measures the distance between the mean intensities in the observation and the particle prediction:

$$d_{\mu}^2(\mathbf{h}_t, I) = \left[\mu_t - \frac{1}{|V(\mathbf{h}_t)|} \sum_{v \in V(\mathbf{h}_t)} I_v \right]^2. \quad (8)$$

We form the likelihood model by combining equations (7) and (8):

$$p(\mathbf{z}_t|\mathbf{h}_t) = \frac{1}{Z} \exp\{-(d_{\nabla}^2 + \lambda d_{\mu}^2)\}, \quad (9)$$

where λ trades off between the two measures and Z is the partition function. In practice, Z does not need to be explicitly computed as the weights (likelihoods) of all samples are normalized at the end of each iteration.

Partial volume effects play a significant role in this computation. Most nerve bundle voxels also contain volume from surrounding anatomy. Modeling large segments of the nerve tracts, as opposed to cross-sections, improves the robustness of the method by increasing the number of voxels that contribute to the distance computations in (7) and (8).

Implementation Details. We initialize each nerve bundle with two nearby clicks that specify $\mathbf{p}_{2,0}$ and $\mathbf{p}_{3,0}$ for a single particle of weight $w = 1$ at time $t = 0$. We set parameters as follows: length $L = 5$, curvature $\sigma_\alpha = 0.4$, radius $\sigma_r = 1$, intensity $\sigma_\mu = 2$. The curvature variation only needs to be increased if sharp transitions are expected. In our experience, the length, radius and intensity parameters rarely need to be varied for similar contrast and resolution MR images. We also consider the intensity distance equally important to the gradient field, setting the tradeoff parameter λ in the likelihood model to 1.

We run the particle filter until all bundles reach the end of the volume. In each iteration, we sample 5,000 particles and keep the ten top-weighted particles to form the sample set $\{\mathbf{h}_t^{(k)}, w_t^{(k)}\}_{k=1}^{10}$. To identify the most appropriate of the surviving tracts, we simulate each tract $V(\mathbf{h}_{1:t})$ in the volume, and compute the likelihood (9) for the whole tract. We retain the top three tracts. Note that due to the multi-hypothesis nature of particle filters, branching nerves are naturally handled. Typical runtime ranges from 10 to 30 CPU-minutes per nerve.

3 Results

We demonstrate our method on MRI scans of the spine in ten nerve bundles from five subjects (3D Wide-band Steady State Free Precession, in-plane resolution of 0.44-0.6mm, slice thickness of 1.2-1.8mm, TR=6.4-6.9ms, TE=2.1-2.4ms, $\Theta=25, \pm 32$ KHz bandwidth, FOV=14 cm) [16]. These include four nerves in two pathologies where the nerves have been displaced by disc herniations. Fig. 3 illustrates an example automatic segmentation. We obtain both expert and automatic segmentations of nerve bundles and ganglia that were deemed traceable, and evaluate tracing accuracy for all bundles. Fig. 3 shows cumulative distributions of the distances between segmentations in terms of centerlines and surfaces.

The mean distance between automatically extracted centerlines and their manual counterparts is within 1 voxel (0.2 ± 0.1 and 1.0 ± 0.3 for automatic-to-manual and manual-to-automatic, respectively); 90% of automatically extracted centerlines are within 0.9 ± 0.3 voxels of the expert centerline, and 90% of expert centerlines are within 2.2 ± 0.5 voxels of the algorithm centerline. Visually, we find very good centerline alignment (≤ 1 voxel) inside the vertebral canal. The greater disagreement is present in the ganglia, where the algorithm tends to under-segment and may be off-center, thereby shifting the centerline by a couple of voxels.

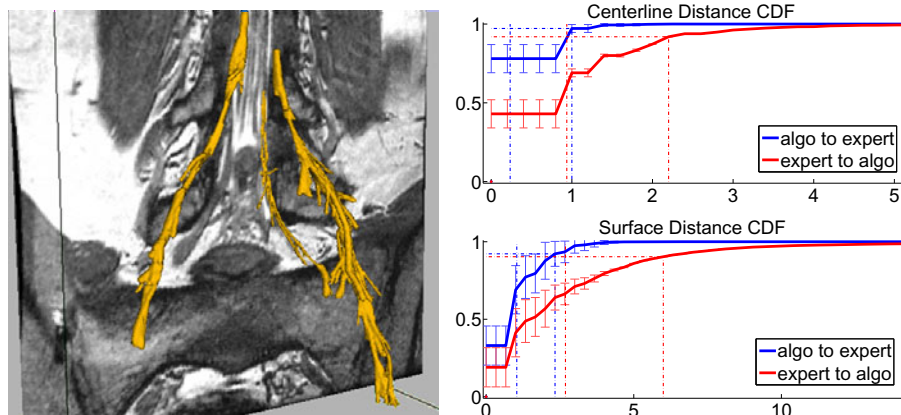


Fig. 3. Left: Example segmentation results – the rightmost nerve shows results without post-processing pruning, while the left segmentation was processed after completion of tracking, as described at the end of Section 2. Right: Cumulative distribution functions of distances (in voxels) between the automatically extracted and manual segmentations: centerline distances on top and surfaces on the bottom. Blue plots indicate differences from algorithm to expert segmentation voxels, red plots correspond to expert-to-algorithm distances. Error bars indicate standard deviations. Dotted vertical lines indicate mean (lower) and 90th percentile distances (higher).

We find that since the edges of nerve bundles and ganglia are subject to the partial volume effects and are often near other anatomical structures, the algorithm slightly over-estimates the extent of the nerves in some regions of the image. As mentioned above, inside the thick ganglia the algorithm under-segments due to the ganglia’s more irregular shape. The mean distance from the automatically extracted surface to the expert surface is 1.1 ± 0.4 voxels and the 90th percentile is at 2.3 ± 0.6 voxels. Visual inspection reveals that the expert segmentation can be irregular and include small deformations or protrusions, especially in pathologies. Since the algorithm attempts to maintain the estimation within the ganglia, the correct outer surface voxels in areas of small irregularities will be more distant from the algorithm prediction. This occurs in cases where the nerve bundle is thicker, generally leading up to and including the ganglia, which can reach 15 voxels in diameter. This results in a mean distance between surfaces of 2.6 ± 0.8 voxels, with the 90th percentile within 5.9 ± 1.9 voxels. We therefore conclude that the proposed segmentation may slightly over-segment (usually by no more than two voxels) in thin areas and under-segment in thick areas, but will give a very good estimation of the nerve core and location.

We also evaluated the algorithm on the nerves following the ganglia, where they split up into several thinner peripheral nerves. Here, the algorithm often loses some nerves due to loss of contrast and the small radii of the nerves. When the algorithm continues to track, we observe a fully estimated path (the segmentation follows the nerve), but with over-segmentation as the filter is driven by stronger edges from the neighboring anatomical structures.

4 Conclusion

We demonstrate tracking of nerve bundles in high-resolution spine MRI. The tracking method is based on particle filtering and requires minimal input from the user. We model nerve segments via Bézier curves and describe a dynamics model for propagating the segments. A new distance measure that utilizes gradient fields and nerve intensities is used to score nerve segments and whole bundles. We demonstrate the method's capacity to handle nerves and ganglia in presence of both high and low contrast. Precisely estimating edges from the current segmentation and segmentation of much thinner peripheral nerves of lower contrast remain challenging problems for future work.

Acknowledgements. This work was supported in part by NAMIC (NIH NIBIB NAMIC U54-EB005149), and the NSF CAREER grant 0642971.

References

1. Aylward, S.R., Bullitt, E.: Initialization, noise, singularities, and scale in height ridge traversal for tubular object centerline extraction. *IEEE Trans. Med. Imag.* 21(2), 61–75 (2002)
2. Balbi, V., Budzik, J.F., Duhamel, A., Bera-Louville, A., Le Thuc, V., Cotten, A.: Tractography of lumbar nerve roots: initial results. *Eur. Radiol* 21, 1153–1159 (2010)
3. Behrens, T., Rohr, K., Stiehl, H.S.: Segmentation of tubular structures in 3d images using a combination of the hough transform and a kalman filter. In: Radig, B., Florczyk, S. (eds.) *DAGM 2001*. LNCS, vol. 2191, pp. 406–413. Springer, Heidelberg (2001)
4. Benmansour, F., Cohen, L.D.: A new interactive method for coronary arteries segmentation based on tubular anisotropy. In: *ISBI 2009*, pp. 41–44 (2009)
5. Boskamp, T., Rinck, D., Link, F., Kummerlen, B., Stamm, G., Mildenerger, P.: New Vessel Analysis Tool for Morphometric Quantification and Visualization of Vessels in CT and MR Imaging Data Sets. *Radiographics* 24(1), 287–297 (2004)
6. Bruijns, J.: Fully-automatic branch labelling of voxel vessel structures. In: *Vision Modeling and Vis.* 2001, pp. 341–350 (2001)
7. Delingette, H., Montagnat, J.: Shape and topology constraints on parametric active contours. *Computer Vision and Image Und.* 83, 140–171 (2000)
8. Florin, C., Paragios, N., Williams, J.: Particle filters, a quasi-monte carlo solution for segmentation of coronaries. In: Duncan, J.S., Gerig, G. (eds.) *MICCAI 2005*. LNCS, vol. 3749, pp. 246–253. Springer, Heidelberg (2005)
9. Fridman, Y., Pizer, S.M., Aylward, S., Bullitt, E.: Segmenting 3d branching tubular structures using cores. In: Ellis, R.E., Peters, T.M. (eds.) *MICCAI 2003*. LNCS, vol. 2879, pp. 570–577. Springer, Heidelberg (2003)
10. Gulsun, M., Tek, H.: Robust vessel tree modeling. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) *MICCAI 2008, Part I*. LNCS, vol. 5241, pp. 602–611. Springer, Heidelberg (2008)
11. Lesagea, D., Angelini, E.D., Bloch, I., Funka-Leaa, G.: A review of 3D vessel lumen segmentation techniques: Models, features and extraction schemes. *Med. Image Anal.* 13(6), 819–845 (2009)

12. Li, H., Yezzi, A.: Vessels as 4-d curves: Global minimal 4-d paths to extract 3-d tubular surfaces and centerlines. *IEEE Trans. Med. Imag.* 26(9), 1213–1223 (2007)
13. Lorigo, L.M., Faugeras, O.D., Grimson, E.L., Keriven, R., Kikinis, R., Nabavi, A., Westin, C.-F.: Curves: Curve evolution for vessel segmentation. *Med. Image Anal.* 5(3), 195–206 (2001)
14. Mille, J., Boné, R., Cohen, L.D.: Region-based 2d deformable generalized cylinder for narrow structures segmentation. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) *ECCV 2008, Part II. LNCS*, vol. 5303, pp. 392–404. Springer, Heidelberg (2008)
15. Schaap, M., Manniesing, R., Smal, I., van Walsum, T., van der Lugt, A., Niessen, W.: Bayesian tracking of tubular structures and its application to carotid arteries in CTA. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) *MICCAI 2007, Part II. LNCS*, vol. 4792, pp. 562–570. Springer, Heidelberg (2007)
16. Schmidt, E.J., Shankaranarayanan, A., Jaume, S., Danagouliau, G., Mukundan, S.J., Nayak, K.S.: Wide-band steady state free precession with small diffusion gradients for spine imaging: Application to superior nerve visualization. In: 18th *ISMRM*, p. 448 (2010)
17. Gordon, N.J., Salmond, D.J., Smith, A.F.M.: Novel approach to nonlinear / non-Gaussian Bayesian state estimation. *IEE Proc-F* 140(2), 107–113 (1993)
18. Bartels, R.H., Beatty, J.C., Barsky, B.A.: *Bézier Curves (Ch. 10): An Introduction to Splines for Use in Computer Graphics and Geometric Modelling*, pp. 211–245. Morgan Kauf., SF (1998)
19. Tyrrell, J., di Tomaso, E., Fuja, D., Tong, R., Kozak, K., Jain, R., Roysam, B.: Robust 3-d modeling of vasculature imagery using superellipsoids. *IEEE Trans. Med. Imag.* 26(2), 223–237 (2007)
20. Yi, J., Ra, J.B.: A locally adaptive region growing algorithm for vascular segmentation. *Int. J. Imag. Syst. Tech.* 13(4), 208–214 (2003)

A Multichannel Markov Random Field Approach for Automated Segmentation of Breast Cancer Tumor in DCE-MRI Data Using Kinetic Observation Model

Ahmed B. Ashraf, Sara Gavenonis, Dania Daye, Carolyn Mies,
Michael Feldman, Mark Rosen, and Despina Kontos

The University of Pennsylvania, Philadelphia, PA 19104, USA
Ahmed.Ashraf@uphs.upenn.edu

Abstract. We present a multichannel extension of Markov random fields (MRFs) for incorporating multiple feature streams in the MRF model. We prove that for making inference queries, any multichannel MRF can be reduced to a single channel MRF provided features in different channels are conditionally independent given the hidden variable. Using this result we incorporate kinetic feature maps derived from breast DCE MRI into the observation model of MRF for tumor segmentation. Our algorithm achieves an ROC AUC of 0.97 for tumor segmentation. We present a comparison against the commonly used approach of fuzzy C-means (FCM) and the more recent method of running FCM on enhancement variance features (FCM-VES). These previous methods give a lower AUC of 0.86 and 0.60 respectively, indicating the superiority of our algorithm. Finally, we investigate the effect of superior segmentation on predicting breast cancer recurrence using kinetic DCE MRI features from the segmented tumor regions. A linear prediction model shows significant prediction improvement when segmenting the tumor using the proposed method, yielding a correlation coefficient $r = 0.78$ ($p < 0.05$) to validated cancer recurrence probabilities, compared to 0.63 and 0.45 when using FCM and FCM-VES respectively.

Keywords: Breast DCE MRI, breast tumor segmentation, tumor characterization, breast cancer recurrence prediction.

1 Introduction

Crucial to the performance of a feature extraction and image classification system is the availability of a reliable segmentation approach for the object of interest (e.g. tumor). In most medical imaging scenarios the automation of this step is particularly important because of the large amount of images to be analyzed. This makes the manual segmentation approach tedious and prohibitively expensive. As a result, a lot of research has been done in the medical imaging community for improving the quality of automated segmentation. Specifically,

there is an abundance of methods in the literature that are geared towards segmenting specific anatomical structures (e.g. parts of the brain). Notable among them are the variants of active contours [1] and active shape models. There are two drawbacks associated with this class of methods. First, they require manual initialization that should be very close to the actual structure to be segmented. Second, they aim at segmenting particular anatomical structures that have well defined control points in their shape, which is often not the case for arbitrarily shaped lesions.

Among the work done on breast tumor segmentation the most popular approach is fuzzy *c*-means (FCM) clustering employed by many researchers due to its simplicity (e.g. [2]). More recently Lee et al. [3] presented a method that runs FCM on the variance map of enhancement kinetics (FCM-VES). These methods do not take into account the overlap of the feature values of the tumor and non-tumor pixels. As a result they have to settle with a manually set threshold on FCM membership probabilities, leading to poor generalization. In order to address these issues we propose to incorporate a component of learning for breast tumor segmentation. We present a multichannel extension of Markov Random Fields (MRFs) to make maximal usage of multiple feature streams derived from the imaging data, here specifically related to kinetic analysis of DCE breast MR images.

In this paper we explore how inference methods like loopy belief propagation [6] may be extended for a multichannel MRF. Zhang et al's work [4] comes closest in concept to our current work. In [4] the authors address the problem of brain image segmentation by making use of conditional independence in MRFs and present an expectation maximization framework to solve the MRF. However, our approach differs fundamentally from [4] in that we exploit conditional independence for solving an MRF via loopy belief propagation [6] that reduces a multichannel MRF to a single channel MRF for inference queries. Moreover, unlike [4], we model our MRF using the superpixel representation that avoids the complexity of dealing with raw pixels. Key contributions of this paper are :

- We prove that for making inference queries, any multi channel MRF can be reduced to a single channel MRF provided the features in different channels are conditionally independent given the hidden variable. (Section 3)
- To elaborate on this premise, we introduce multiple feature channels derived from the kinetic analysis (pixel wise feature maps of enhancement and wash-in-slope) of breast dynamic contrast enhanced (DCE) magnetic resonance (MR) images in the observation model of MRF. (Section 4)
- We show that the segmentation algorithm based on our proposed method yields an AUC of 0.97 under the ROC curve for breast tumor segmentation compared to 0.86 and 0.60 for commonly used previously proposed approaches, [2], [3]. (Section 5)
- Finally we demonstrate that superior segmentation leads to improved breast tumor characterization. We show that the segmentation methods of this paper lead to improvement in predicting validated breast cancer recurrence scores from DCE breast MR images. (Section 6)

We begin with a brief review of single channel Markov random fields.

2 Single Channel MRF

To elaborate on the MRF concept, Figure 1(a) shows how an image can be modeled as a Markov random field. Each node represents the class of a pixel, and neighboring pixels are connected via edges. In the context of segmentation the goal is to infer the class label for each pixel (e.g. foreground vs. background). Throughout the paper we choose the super pixel representation [5] as its computational complexity is much less than the pixel representation. Each node emits an observation which can be any feature e.g. pixel value. The joint probability of super pixel class and feature value over the entire image is given in Eq. 1:

$$P(\mathcal{X}, \mathcal{Y}) \propto \prod_i \phi_i(x_i) \prod_{i,j \in E} \psi_{ij}(x_i, x_j). \quad (1)$$

$$\phi_i(x_i = c) = P(y_i | \mathcal{GMM}_c). \quad (2)$$

$$\psi_{ij}(x_i, x_j) = e^{-\beta \mathcal{I}(x_i \neq x_j)}. \quad (3)$$

In Equation 1, ϕ represents the node-potential and ψ represents the edge potential, and E models the adjacency of nodes including only those nodes that have edges between them. $\phi_i(x_i)$ captures the correlation between the feature and the class label, indicating the likelihood of x_i coming from class c based on the feature value of y_i and the feature distribution of class c . Node potentials can be learned from the training data. In Equation 2, \mathcal{GMM}_c is a Gaussian mixture model learned for class c . In Equation 3, ψ biases neighboring nodes to have the same class label via the parameter β . Approximate MAP (maximum a posteriori) solution to the MRF can be inferred by loopy belief propagation (Loopy BP) [6] which maximizes the joint probability $P(\mathcal{X}, \mathcal{Y})$. Loopy BP is a dynamic message passing algorithm used for doing inference in MRF. $\delta_{i \rightarrow j}(\mathcal{X}_j)$ is defined to be an incoming message into node j from its neighboring node i . To start the process of message passing through the nodes, we need to initialize the node messages (typically to unity) and then the messages can be updated in the next iteration as follows:

$$\delta_{i \rightarrow j}(\mathcal{X}_j) = \frac{1}{Z_{i \rightarrow j}} \sum_i \phi_i(x_i) \cdot \psi_{ij}(x_i, x_j) \prod_{k \in \mathcal{N}(i) - j} \delta_{k \rightarrow i}(x_i). \quad (4)$$

where $Z_{i \rightarrow j}$ is a normalization constant, and $\mathcal{N}(i)$ is a set containing the neighbors of node i . Equation 4 is repeatedly invoked till the messages converge (the update in each message is less than ϵ e.g. 10^{-6}). Once the messages have converged the final inference is done by using Equation 5 below:

$$\hat{P}(\mathcal{X}_i) = \frac{1}{Z_i} \phi_i(x_i) \prod_{k \in \mathcal{N}(i)} \delta_{k \rightarrow i}(\mathcal{X}_i). \quad (5)$$

The inference engine will output for each node a vector of size $C \times 1$ ($C = 2$ for two classes), representing the belief of this node coming from each class. Optimal class label is simply the class with the highest belief.

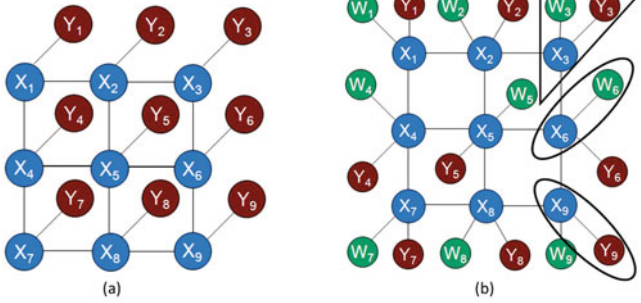


Fig. 1. (a) A single channel MRF: each node represents a super pixel emitting an observation (see Section 2) (b) A multichannel model with two observations . The variables inscribed in the triangle represent the joint distribution of the MRF model. By invoking Theorem 1 we can factorize that joint distribution into two parts represented by the ellipses. (see Section 3)

3 Multichannel MRF

Building on the single channel MRF presented in the last section we now discuss how to incorporate multiple observations in the MRF model. To elaborate on the concept, Figure 1(b) illustrates a multichannel MRF that includes two observations (similar notion could be extended to more than two observations). The joint probability of the super pixel class and the two observations over the entire image is given below:

$$P(\mathcal{X}, \mathcal{Y}, \mathcal{W}) \propto \prod_i \phi_{i(Multichannel)}(x_i, y_i, w_i) \prod_{i,j \in E} \psi_{ij}(x_i, x_j) . \quad (6)$$

$$\phi_{i(Multichannel)}(x_i = c, y_i, w_i) = P(y_i, w_i | \mathcal{GMM}_c) . \quad (7)$$

where $\phi_{i(Multichannel)}$ is the multichannel node potential for the two observations. Our main premise here is that the machinery presented in the previous section can be extended for solving the above multichannel MRF. To this end, we present the following theorem.

Theorem 1: Any multichannel MRF can be reduced to a single channel MRF for making inference queries if the features in different channels are conditionally independent given the hidden (inferred) variable.

Proof: We prove the above theorem for a two channel MRF, while the same notion can be extended to more than two channels without loss of generality. If the features in different channels are conditionally independent given the hidden variable we can factorize the right hand side of Equation 7 as follows:

$$P(y_i, w_i | \mathcal{GMM}_c) = P(y_i | \mathcal{GMM}_c) P(w_i | \mathcal{GMM}_c) \text{ if } Y \perp W | X . \quad (8)$$

Equivalently Equation 8 can be written as:

$$\phi_{i(Multichannel)}(x_i = c, y_i, w_i) = \phi_i^{(y)}(x_i = c, y_i) \cdot \phi_i^{(w)}(x_i = c, w_i) . \quad (9)$$

where $\phi_i^{(y)}$ and $\phi_i^{(w)}$ are the individual node potentials for the features y and w respectively. We can now substitute these individual node potentials in the message update and inference equations as follows:

$$\delta_{i \rightarrow j}(\mathcal{X}_j) = \frac{1}{Z_{i \rightarrow j}} \sum_i \underbrace{\phi_i^{(y)}(x_i = c, y_i) \cdot \phi_i^{(w)}(x_i = c, w_i)} \cdot \psi_{ij}(x_i, x_j) \prod_{k \in \mathcal{N}(i) - j} \delta_{k \rightarrow i}(x_i) .$$

$$\hat{\mathcal{P}}(\mathcal{X}_i) = \frac{1}{Z_i} \underbrace{\phi_i^{(y)}(x_i = c, y_i) \cdot \phi_i^{(w)}(x_i = c, w_i)} \prod_{k \in \mathcal{N}(i)} \delta_{k \rightarrow i}(\mathcal{X}_i) . \quad (10)$$

The individual node potentials in the above equations (the underlined portion) can be pre-computed and be combined as a single potential. As a consequence, the inference engine would now be running queries on a multichannel MRF without incurring any additional cost, effectively reducing it to a single channel MRF, proving the claim of Theorem 1 \square

In the next section we elaborate on the extraction of pixel-wise feature maps to build a kinetic observation model for our multichannel MRF.

4 Kinetic Observation Model

Typically, for DCE MR images we have a pre-contrast image (captured prior to the injection of a contrast agent) and a number of post contrast images, captured at different time points after the injection of the contrast agent. A usual way to quantify the enhancement pattern [3] is to compute percentage enhancements relative to the pre-contrast image. If we compute the relative enhancement on a pixel by pixel basis we can achieve pixel wise maps of the contrast enhancement. For a particular pixel, the enhancement plotted as a function of time is termed as the kinetic curve. In the literature (e.g. [3]) a number of basic features can be computed from this kinetic curve. For example, peak enhancement (PE), time to peak (TTP), wash-in-slope (WIS), wash-out-slope (WOS). Figure 2(a) illustrates

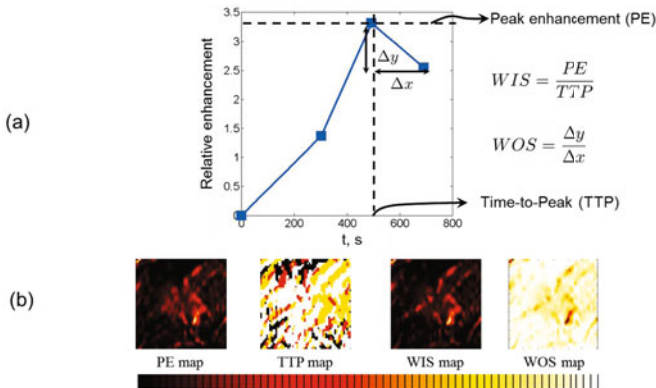


Fig. 2. (a) Illustration of basic kinetic features for a single pixel. (b) Pixel wise maps.

these features for a single pixel. However, we can derive a rich kinetic feature set by computing the pixel-wise map for each feature as depicted in Figure 2(b). We aim to investigate the utility of these feature maps if included as an observation model in our multichannel MRF for the purpose of breast tumor segmentation.

5 Segmentation Experiments

Dataset: All experiments presented in this paper were conducted on DCE breast MR images of 60 women diagnosed with breast cancer. These women had estrogen receptor positive (ER+), node negative tumors, which were analyzed with the Oncotype DX prognostic gene expression assay. Oncotype DX is a validated reverse-transcriptase-polymerase-chain-reaction (RT-PCR) assay (developed by Genomic Health Inc.) that measures the expression of 21 genes in RNA from formalin-fixed paraffin-embedded (FFPE) tumor tissue samples from the primary breast cancer [7]. The final outcome of the Oncotype DX assay is a continuous recurrence score that predicts the likelihood of breast cancer recurrence in 10 years after the treatment. To learn feature statistics for distinguishing the breast tumor and non-tumor area of an image, a fellowship-trained board-certified breast imaging radiologist delineated the lesion boundaries.

Feature selection: To select the features that meet the conditions of Theorem 1 we need to assess the conditional independence of features for which we use the mutual information as a criterion. In the multichannel MRF of Figure 1(b), we are interested in assessing the independence of Y and W given X . Their conditional mutual information is given by:

$$I(Y; W|X) = \sum_{x,y,w} p(x, y, w) \log \frac{p(y, w|x)}{p(y|x)p(w|x)}. \quad (11)$$

We computed $I(Y; W|X)$ by sequentially setting (Y, W) to all possible pairs of the feature maps (post contrast enhancement, peak enhancement, wash-in-slope, wash-out-slope), and retaining the pairs for which $I(Y; W|X) \leq thresh$ (0.02). The feature pair with the minimum $I(Y; W|X)$ was then selected. The above process resulted in the selection of the following two feature maps for our MRF: first post-contrast relative enhancement and the wash-in-slope.

To enable maximal usage of the training data we used a leave-one-out cross validation strategy. To reduce the complexity of the multichannel MRF, we first over-segmented the DCE MR images using the super pixels approach based on normalized graph cuts [5]. In order to define the MRF neighborhood for superpixels we scan the rows and columns of the superpixelized image and look for transitions. This enables to build the adjacency matrix for superpixels that captures the neighborhood. In Figure 3, we show representative segmentation results of our algorithm (row 5). For comparison, we also show segmentation results for FCM as used in [2] (row 3), and results based on running FCM on enhancement variance FCM-VES [3], (row 4). A quantitative comparison in terms of ROC's for the three segmentation strategies is given in Figure 4(a). Our proposed method yields an AUC of 0.97 under the ROC curve as compared to 0.86 and 0.60 for [2] and [3] respectively.

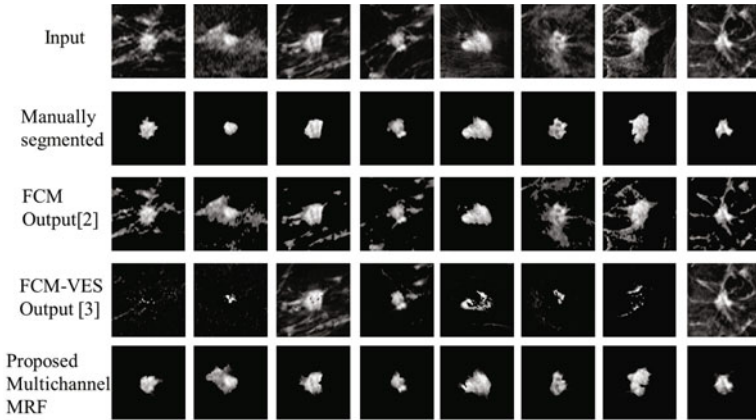


Fig. 3. Qualitative comparison of segmentation results. Results of multichannel MRF approach (Row 5) are qualitatively more similar to the ground truth (Row 2) as compared to traditional approaches (Row 3,4).

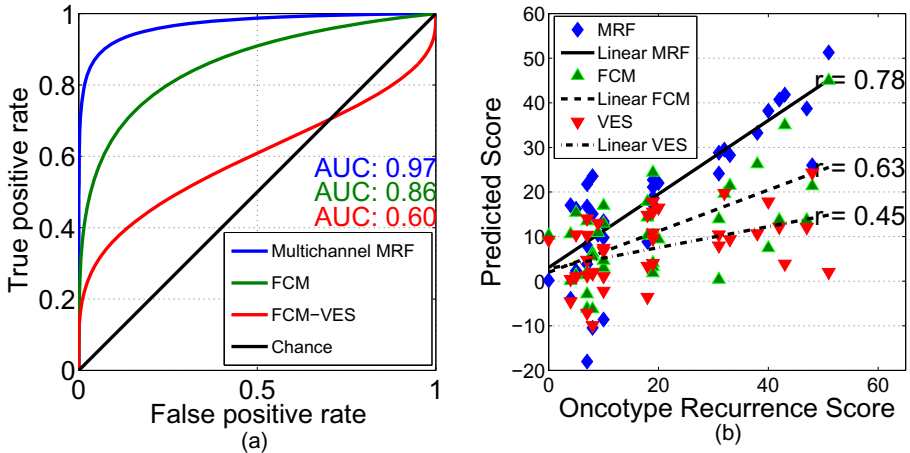


Fig. 4. (a) ROC comparison for breast tumor segmentation. (b) Comparison of Oncotype DX score prediction.

6 Oncotype Score Prediction Experiments

In this section we investigate if better quality segmentation leads to a better tumor characterization. To this end, we aim to predict the Oncotype DX recurrence scores from DCE breast MR images based on previously validated kinetic features [2] extracted within the tumor region segmented as the outcome of our segmentation algorithm. Briefly, in order to extract features from the segmented tumor we cluster the pixels on the basis of their time to peak. In the DCE MR images used in this study there were three post contrast time points. We thus partition the pixels into three groups based on the similarity of their kinetic behavior. Using this partition we first compute the percentage of pixels in each group (this

gives us three features). Within each group we then compute statistics (up to second order) for each of the following maps: peak enhancement, wash-in-slope, wash-out-slope [2], [3]. For every feature map there are three groups and two features per group (mean and variance). In all we build a vector of 21 simple kinetic features. Using these features we learn a linear regression model to predict the continuous Oncotype DX scores. To maximize the utility of the training data we employed a leave-one-out cross validation strategy. The results of the prediction models based on each of the segmentation strategies (multichannel MRF, FCM, FCM-VES) are presented in Figure 4(b). We obtain a correlation coefficient of $r = 0.78$ to the Oncotype DX cancer recurrence scores using our segmentation method as compared to correlations of 0.63 and 0.45 for the other segmentation methods, indicating that our proposed improved segmentation can also result to superior tumor characterization.

7 Concluding Remarks

In this paper we have presented a framework for incorporating multiple feature streams in Markov random fields. We have shown how a multichannel MRF can be reduced effectively to a single channel MRF from the perspective of inference queries. Using this framework we incorporate a kinetic observation model derived from DCE breast MR images into the MRF and demonstrate superior segmentation results as compared to previous methods. Moreover, we also demonstrate that our improved segmentation leads to better tumor characterization. With this framework in place, future research will also investigate the introduction of multi-modality image sequences in the observation model of MRF.

References

1. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. *International Journal of Computer Vision* 1(4), 321–331 (1988)
2. Bhooshan, N., Giger, M.L., Jansen, S.A., Li, H., Lan, L., Newstead, G.M.: Cancerous breast lesions on DCE MR Images: Computerized characterization for Image based prognostic markers. *Radiology* 254(3), 680–690 (2010)
3. Lee, S.H., Kim, J.H., Cho, N., Park, J.S., Yang, Z., Jung, Y.S., Moon, W.K.: Multilevel analysis of spatiotemporal association features for differentiation of tumor enhancement patterns in breast DCE-MRI. *Med. Phys.* 37, 3940 (2010)
4. Zhang, Y., Brady, M., Smith, S.: Segmentation of Brain MR Images Through a Hidden Markov Random Field Model and the Expectation-Maximization Algorithm. *IEEE Trans. Med. Imag.* 20, 45–57 (2001)
5. Ren, X., Malik, J.: Learning a classification model for segmentation. In: *ICCV*, vol. 1, pp. 10–17 (2003)
6. Meltzer, T., Globerson, A., Weiss, Y.: Convergent message passing algorithms – a unifying view. In: *UAI*, pp. 393–401 (2009)
7. Paik, S., Tang, G., Shak, S., Kim, C., Baker, J., Kim, W., Cronin, M., Baehner, F.L., Watson, D., Bryant, J., Costantino, J.P., Geyer, C.E., Wickerham, D.L., Wolmark, N.: Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. *J. Clin. Oncol.* 24(23), 3726–3734 (2006)

3-D Graph Cut Segmentation with Riemannian Metrics to Avoid the Shrinking Problem

Shouhei Hanaoka^{1,2}, Karl Fritscher¹, Martin Welk¹, Mitsutaka Nemoto², Yoshitaka Masutani², Naoto Hayashi³, Kuni Ohtomo², and Rainer Schubert¹

¹ The Institute of Biomedical Image Analysis, The Health and Life Sciences University (UMIT), Eduard-Wallnöfer-Zentrum 1, Hall in Tirol, Austria

² Dept. of Radiology and

³ Dept. of Computational Diagnostic Radiology and Preventive Medicine, The Univ. of Tokyo Hospital, 7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan
hanaoka-tky@umin.ac.jp

Abstract. Though graph cut based segmentation is a widely-used technique, it is known that segmentation of a thin, elongated structure is challenging due to the “shrinking problem”. On the other hand, many segmentation targets in medical image analysis have such thin structures. Therefore, the conventional graph cut method is not suitable to be applied to them. In this study, we developed a graph cut segmentation method with novel Riemannian metrics. The Riemannian metrics are determined from the given “initial contour,” so that any level-set surface of the distance transformation of the contour has the same surface area in the Riemannian space. This will ensure that any shape similar to the initial contour will not be affected by the shrinking problem. The method was evaluated with clinical CT datasets and showed a fair result in segmenting vertebral bones.

Keywords: Graph cut, Segmentation, Riemannian geometry, Spine.

1 Introduction

Graph cuts are one of the most widely used techniques for segmentation tasks in image analysis. The biggest advantage of the algorithm is that it can solve a segmentation problem as a global optimization problem without iterative calculation, and it guarantees a globally optimal solution [1]. The typical cost function minimized by the algorithm consists of 2 terms; a spatial coherency term and a data term, which are defined as follows:

$$E(f) = \sum_{p,q \in N} V_{p,q}(f_p, f_q) + \sum_{p \in P} D_p(f_p) \quad (1)$$

where D_p is the data energy, $V_{p,q}$ is the smoothness energy, N is the set of neighborhood pairs, f_p is the label assigned to the pixel p , and P are all pixels in the image [2]. Here, the 1st term (spatial coherency term) can be considered as a term which evaluates the length (in 2-D) or area (in 3-D) of the boundary, which is modulated with the

contrast in the image. Therefore, minimizing the energy using this term causes a bias towards shorter boundaries. This behavior is known as “shrinking bias” [3]. It causes severe problems especially when the target object has a long, thin spine-like process. Unfortunately, elongated structures are very common in anatomical objects. This fact significantly aggravates the segmentation of e.g. vertebrae in CT images and causes significant segmentation errors [4]. Though a number of methods have been reported to address the problem, such as [3] and [5], to the best of our knowledge, no simple way to avoid it has been presented.

In 2003, Boykov and Kolmogorov reported a method to construct a graph where cut metric approximates any given Riemannian metric, and utilized it for image segmentation [6]. The metrics were defined and calculated from the gradient information of the image to be segmented.

In this paper, we propose a novel method to avoid the shrinking problem by performing graph cuts in a Riemannian space. The Riemannian metrics are not calculated using image information but using a predefined shape template, or, “initial contour.” Although the position and pose information of the target object is needed in advance, no other prior information like any “seed region” is required.

The basic idea is to compose a 3-D Riemannian space in which any surfaces parallel to the initial contour (an isosurface) has the same surface area. By performing graph cut segmentation in this 3-D Riemannian space, the inside and outside of the object can be considered and handled in the same way, so that the spatial coherency term serves as an evaluator of how the segmentation result differs from its closest isosurface. Therefore, a “shrinking problem” in the usual sense cannot occur.

2 Methods

The basic notion of this segmentation method is to perform graph cut-based segmentation in a Riemannian space which satisfies the following conditions:

- 1) It can be defined everywhere in the input image (excluding the points where the distance transformation of the initial contour, $dist(\mathbf{x})$, is not differentiable).
- 2) Any isosurface (level set surface) of $dist(\mathbf{x})$ — i.e. the set of points \mathbf{x} with $dist(\mathbf{x}) = const$ — has the same surface area.

The “initial contour” approximates the object to be segmented (e.g. the mean shape of the target object). It is assumed that the given grayscale image and the initial contour is registered rigidly in advance. Then, the metrics of the Riemannian space are determined and calculated from the initial contour, so that any isosurface becomes parallel to each other and has the same surface area. Although such a Riemannian space has some singular points with indefinite metrics, our method can be performed in a stable manner even with them.

2.1 Calculation of the Riemannian Metrics

In order to apply the graph cut algorithm in Riemannian space, the Riemannian metrics at every grid point must be determined in advance. In this study, the metrics were calculated from the distance map of the initial contour $dist(\mathbf{x})$.

The metric tensor \mathbf{G} at any point on the initial contour is defined to be isometric, thus, equal to an identity matrix. At any other point, the metrics are defined so that the sum of the area of any iso-surface (any level set of the distance map $dist(\mathbf{x}) = d$) will be identical. More detail on the calculation of the metrics will be given in Sections 2.1.1 and 2.1.2.

2.1.1 Metrics Calculation from Curvatures of the Distance Function

Fig. 1 illustrates the basic idea of the metrics calculation (in a 2-D space for explanatory usage.) Suppose that the metrics at the point \mathbf{x}_d is to be calculated. Let the (outside-positive) signed distance of \mathbf{x}_d from the initial contour be d .

$$dist(\mathbf{x}_d) = d \tag{2}$$

So, there must be a point \mathbf{x}_0 , on the initial contour, whose distance from \mathbf{x}_d equals d .

Let the initial surface (a curve for this 2-D example) be S_0 , and the isosurface including \mathbf{x}_d be S_d .

$$S_0 = \{\mathbf{x} \mid dist(\mathbf{x}) = 0\}, \quad S_d = \{\mathbf{x} \mid dist(\mathbf{x}) = d\} \tag{3}$$

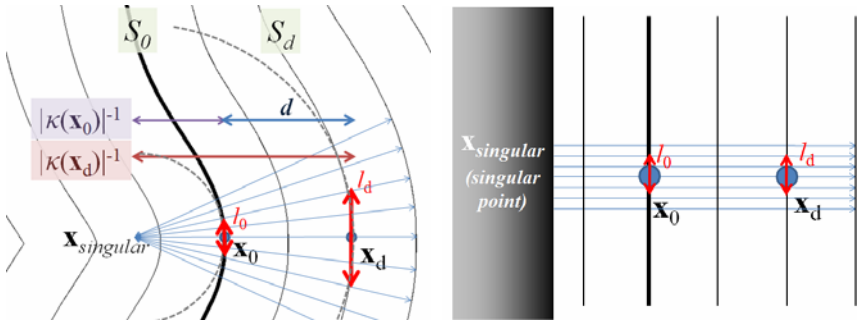


Fig. 1. Determination of the metrics. (left) The line segment l_0 and l_d are determined from the curvature radii, κ_0^{-1} and κ_d^{-1} , of the distance map. (right) Corresponding rectified presentation of the Riemannian space, in which the lengths of the two line segments are identical.

Let the curvature of the curve S_0 at \mathbf{x}_0 be κ_0 , and that of S_d at \mathbf{x}_d be κ_d (see also Fig. 1). Here, the curvatures are defined to be negative when the curve is convex. Under the assumption that both curvatures are negative (or both positive), there must be a point $\mathbf{x}_{singular}$ where the curvature becomes infinite. Note that both curves S_0 and S_d can be locally approximated as arcs whose center point is $\mathbf{x}_{singular}$, as shown in Fig. 1. Therefore, the curvature radii (the inverses of the curvatures) at \mathbf{x}_0 and \mathbf{x}_d equal their distances from $\mathbf{x}_{singular}$. Thus, the curvatures satisfy the following equation:

$$-\kappa_0^{-1} + d = -\kappa_d^{-1} \tag{4}$$

Suppose that there are 2 minimal line segments, l_0 and l_d , which are parallel to the isosurfaces. Let further the Euclidean lengths of the two segments be l_0^E and l_d^E and

the proportion of them be identical to the proportion of the curvature radii κ_0^{-1} and κ_d^{-1} . Using Formula (4), the proportion is calculated as:

$$\frac{l_d^E}{l_0^E} = \frac{\kappa_d^{-1}}{\kappa_0^{-1}} = \frac{1}{1 + \kappa_d \cdot d} \tag{5}$$

In order to achieve the equalization of all iso-surface areas, the lengths of l_0 and l_d in a sense of Riemannian geometry, l_0^R and l_d^R , must be identical (Fig. 1, right). According to the Riemannian geometry theory, the length of a minimal line segment l can be approximated as:

$$length(l) \cong \sqrt{\mathbf{v}^t \mathbf{G} \mathbf{v}} \tag{6}$$

where the length and direction of the vector \mathbf{v} are those of the minimal line l , and a symmetric matrix \mathbf{G} is the Riemannian metric tensor at that point. Furthermore, we have defined the metrics on the initial contour as Euclidean metrics, so that the metrics tensor on the initial contour, \mathbf{G}_0 , is the identity matrix \mathbf{I} . Thus, the length l_0^R is equal to its Euclidean equivalent l_0^E . On the other hand, given the metric tensor \mathbf{G}_d at the point \mathbf{x}_d , the length l_d^R is calculated as:

$$l_d^R \cong l_d^E \cdot \sqrt{\mathbf{w}^t \mathbf{G}_d \mathbf{w}} \tag{7}$$

where \mathbf{w} is a unit vector with the same direction as l_d .

Consequently, the metric tensor \mathbf{G}_d must satisfy the formula below:

$$\mathbf{w}^t \mathbf{G}_d \mathbf{w} = (1 + \kappa_d \cdot d)^2 \tag{8}$$

To satisfy (8), we defined the tensor \mathbf{G}_d as follows:

$$\mathbf{G}_d = \mathbf{R} \cdot \text{diag}((1 + \kappa_d \cdot d)^2, 1) \cdot \mathbf{R}^t \tag{9}$$

$$\mathbf{R} = (\mathbf{w} \quad \mathbf{n}), \quad |\mathbf{w}| = |\mathbf{n}| = 1, \quad \mathbf{w} \cdot \mathbf{n} = 0, \quad \mathbf{n} = \frac{\nabla \text{dist}(\mathbf{x})}{|\nabla \text{dist}(\mathbf{x})|}$$

Here, \mathbf{R} is an orthogonal matrix whose column vectors \mathbf{w} and \mathbf{n} are unit vectors perpendicular and parallel to the gradient vector of the distance map, respectively.

2.1.2 Metrics Calculation for 3-D Volumes

In order to extend the metric calculation to a 3-D space, the metrics must be calculated from two principal curvatures of the given isosurface, instead of only one curvature in 2-D. When the 1st and 2nd curvatures are given as κ_1 and κ_2 , and the corresponding principal directions are given as unit vectors w_1 and w_2 , the metric tensor can be defined as:

$$\mathbf{G}_d = \mathbf{R} \cdot \text{diag}((1 + \kappa_1 \cdot d)^2, (1 + \kappa_2 \cdot d)^2, 1) \cdot \mathbf{R}^t \tag{10}$$

$$\mathbf{R} = (\mathbf{w}_1 \quad \mathbf{w}_2 \quad \mathbf{n}), \quad \mathbf{n} = \frac{\nabla \text{dist}(\mathbf{x})}{|\nabla \text{dist}(\mathbf{x})|}$$

Using these metrics, the areas of all isosurfaces become equal. (The proof is omitted due to space limiting.)

2.2 Modification of the Edge Weights

In order to perform graph cut in the Riemannian space, both the data term and the spatial coherency term must be adequately modified. The former can be defined by the graph's edge weights between each image grid point and the s - (source) or t - (sink) node. The latter is defined by the edge weights between 2 adjacent grid points.

The theoretical framework to perform graph cut with Riemannian metrics was firstly presented by Boykov et al [6] based on integral geometry. However, their original approach is for finding the minimal surface in a Riemannian space, without considering any apparent equivalent of the data term (aside from hard constraints for seed regions). Therefore, we chose another, much simpler solution to modify the terms under the assumptions that:

- 1) The weights of edges between adjacent image grid points have to be proportional to the intersectional area of their border (in the Riemannian space).
- 2) The weights of edges to the s/t node have to be proportional to the volume of the space occupied by the grid point (again in the Riemannian space).

The first assumption is derived from the fact that these weights compose the spatial coherency term, which minimizes the surface area. The second assumption is determined in order to apply the same s/t weight to any unit volume.

In determining the weights, there is a difficulty due to singularity of the distance map. Our definition of metrics depends completely upon differentiability of the distance map. However, it is not differentiable at any point which has multiple nearest points on the initial contour. In our model, any singular point can be considered having infinite metrics, so that it has infinite volume (i.e., an infinite s/t -edge weight). Moreover, the finite difference approximation for the differentials is problematic near the singular points.

To avoid this problem, we restrict the area of interest to a band-like region $|\text{dist}(\mathbf{x})| < d_{\max}$. The constant d_{\max} is determined as the maximal expected dislocation between the initial contour and the true contour to be segmented. The restriction is performed as follows:

- 1) Grid points close to any singular point, or *singular-including* points, are detected.
- 2) All *singular-including* grid points are removed from the graph, as well as any grid points with $|\text{dist}(\mathbf{x})| \geq d_{\max}$.
- 3) Grid points *adjacent to singular* (= adjacent to any *singular-including* point) have to be treated specifically: these grid points are considered to have finite 'depth' d_{\max} towards the singular point (Fig. 2).

At each grid points, the data term is multiplied by the volume occupied by the point in the Riemannian space. Given the grid size is δ , and the metric tensor at the grid center point is \mathbf{G} , the volume occupied by the grid point can be approximated as

$$V = \delta^3 \cdot \sqrt{\det \mathbf{G}} \tag{11}$$

If the grid point is *adjacent to singular*, the occupied volume is considered as the volume within the proximal border of the grid and the border defined by the pre-defined depth d_{max} . It is simply approximated by multiplying the volume V by $(d_{max} - |dist(\mathbf{x})|)$.

The non-s/t edge weights are modified by multiplying the cross-sectional area A . A is approximated as follows:

$$A = \delta^2 \cdot \sqrt{(\mathbf{u}_1^t \cdot \mathbf{G} \cdot \mathbf{u}_1) \cdot (\mathbf{u}_2^t \cdot \mathbf{G} \cdot \mathbf{u}_2) - (\mathbf{u}_1^t \cdot \mathbf{G} \cdot \mathbf{u}_2)^2} \tag{12}$$

where \mathbf{u}_1 and \mathbf{u}_2 are unit vectors perpendicular to the line segment connecting the two grid points and perpendicular to each other.

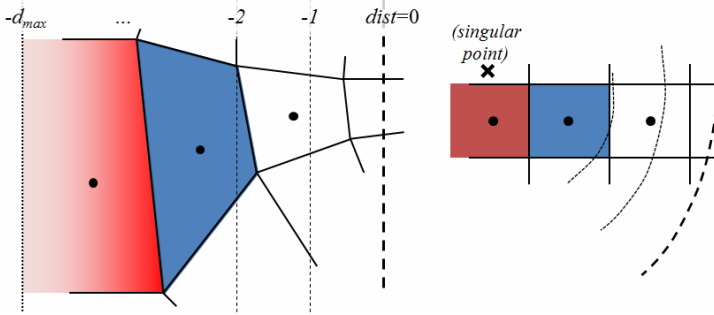


Fig. 2. Schematic views of image grids, (left) in a rectified presentation of the Riemannian space, and (right) in the original image space. The two painted boxes represent the areas occupied by two grid points. The left grid point is referred to as *adjacent to singular*.

2.3 Evaluation

The method was evaluated on 220 human vertebral bones in 10 datasets of clinical computed tomography (CT) images. The 1st and 2nd cervical vertebrae were excluded because of their unique shapes. The vertebral bones were divided into 4 groups: cervical, upper thoracic, lower thoracic and lumbar ones. In each group, the mean shape was calculated and used as initial contour. Before the proposed method was applied, each vertebra was cropped and rigidly registered in the same manner as described in [4]. Though 209 vertebrae were correctly identified and cropped by this full automatic process, 11 were failed and manually corrected in this study. Therefore, the following experiments were performed using these 220 pose-compensated, cropped volume images of vertebrae. The voxel size of $1 \times 1 \times 1$ mm and $d_{max}=24$ mm were used in this study.

The data term used in this study (before modified with metrics) is a binary function whose value is -1 and 1 within and out of the initial contour, respectively. The spatial coherency term is a modified version of frequently used term introduced by Boykov et al. in [7]. In the modification only inward-positive image gradients are considered and outward-positive ones are regarded as 0 to detect outline of the bones.

The similarity index, Hausdorff distance and mean distance compared to the manual segmentation results were calculated for each vertebra. Each result was compared to that of a non-Riemannian version of our graph cut segmentation method. It was initialized with the same parameters and data terms as the Riemannian graph cut version without any Riemannian metric modification.

3 Experimental Results and Discussions

The results of vertebral bone segmentation are shown in Table 1. An example result is shown in Fig. 3. The results of the proposed method turned out to be superior to the conventional graph cut approach in all criteria and in all 4 vertebral groups. The overall mean distance error (\pm s. d.) of the proposed method and the conventional method was 1.28 ± 0.65 mm and 3.76 ± 2.67 mm, respectively.

The result was comparable to another method reported in our previous work [4] based on shape-intensity combined statistical models, in which the overall mean distance was 1.28 ± 1.52 mm. It was also comparable to the study reported by Klinder et al. [8], in which the overall mean distance was 1.12 ± 1.04 mm. On the other hand, the proposed method was less precise in thoracic vertebrae, mainly due to incorrect segmentation of the region adjacent to the ribs. For thoracic vertebrae, another study by Ma et al. [9] reported a better result with 0.95 ± 0.91 mm of the mean distance.

One of limitations of the proposed method is that the initial contour must be given in advance. Therefore, it is possible that the method is less effective for targets with larger shape variation. On the contrary, it might be especially effective for objects with less variable but more complex shapes (e.g., with many thin or protruding parts), because most of known segmentation methods are not good at segmenting such complex objects reliably.

Table 1. The segmentation results. (mean \pm s.d.)

Similarity index	Cervical	Upper th.	Lower th.	Lumbar
with proposed metrics	0.77 \pm 0.03	0.79 \pm 0.06	0.86 \pm 0.03	0.87 \pm 0.04
with Euclidean metrics	0.45 \pm 0.21	0.48 \pm 0.27	0.68 \pm 0.23	0.82 \pm 0.05
Hausdorff dist. (mm)	Cervical	Upper th.	Lower th.	Lumbar
with proposed metrics	9.03 \pm 2.15	17.10 \pm 7.71	11.13 \pm 5.61	11.80 \pm 4.26
with Euclidean metrics	17.62 \pm 4.16	22.08 \pm 6.84	18.05 \pm 7.34	16.93 \pm 5.07
Mean dist. (mm)	Cervical	Upper th.	Lower th.	Lumbar
with proposed metrics	1.11 \pm 0.24	1.75 \pm 0.90	1.11 \pm 0.51	1.11 \pm 0.40
with Euclidean metrics	4.16 \pm 2.00	4.90 \pm 3.1	3.53 \pm 2.89	2.14 \pm 1.01



Fig. 3. An example of segmentation result for a cervical vertebra. (From left) The original grayscale image, rigidly-registered initial contour and two graph cut results with Euclidean and proposed metrics, respectively.

4 Conclusion

An approach of graph cut segmentation based on a newly introduced Riemannian metrics was presented. The experimental result suggested its advantage in segmenting thin, spine-like structures in which conventional graph cut methods are affected by the “shrinking problem.” Despite the simplicity of the approach, it achieves in some situations even comparable segmentation quality as more complex model-based methods. In future work, we aim at integrating the Riemannian graph cut and model-based approaches in order to develop a more powerful and accurate segmentation scheme.

Acknowledgement. This study is a part of the research project “Computational Anatomy for Computer-aided Diagnosis and therapy: Frontiers of Medical Image Sciences”, which is financially supported by the grant-in-aid for scientific research on innovative areas MEXT, Japan.

References

1. Peng, B., Veksler, O.: Parameter selection for graph cut based image segmentation. In: Proc. of British Machine Vis. Conf. (2008)
2. Vineet, V., Narayanan, P.: CudaCuts: Fast graph cuts on the GPU. In: CVPR 2008 Workshop on Visual Computer Vision on GPUs (CVGPU 2008), pp. 1–8 (2008)
3. Vicente, S., Kolmogorov, V., Rother, C.: Graph cut based image segmentation with connectivity priors. In: CVPR (2008)
4. Hanaoka, S., Fritscher, K.D., Schuler, B., Masutani, Y., Hayashi, N., Ohtomo, K., Schubert, R.: Whole vertebral bone segmentation method with a statistical intensity-shape model based approach. In: SPIE Medical Imaging (2011)
5. Kolmogorov, V., Boykov, Y.: What metrics can be approximated by Geo-Cuts, or global optimization of length/area and flux. In: ICCV, vol. I, pp. 564–571 (2005)
6. Boykov, Y., Kolmogorov, V.: Computing geodesics and minimal surfaces via graph cuts. In: ICCV, vol. I, pp. 26–33 (2003)
7. Boykov, Y., Jolly, M.-P.: Interactive graph cuts for optimal boundary and region segmentation of objects in N-D images. In: ICCV, vol. I, pp. 105–112 (2001)
8. Klinder, T., Ostermann, J., Ehm, M., Franz, A., Kneser, R., Lorenz, C.: Automated model-based vertebra detection, identification, and segmentation in CT images. *Medical Image Analysis* 13(3), 471–482 (2009)
9. Ma, J., Lu, L., Zhan, Y., Zhou, X., Salganicoff, M., Krishnan, A.: Hierarchical Segmentation and Identification of Thoracic Vertebra Using Learning-Based Edge Detection and Coarse-to-Fine Deformable Model. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part I. LNCS, vol. 6361, pp. 19–27. Springer, Heidelberg (2010)

Adaptive Multi-cluster Fuzzy C-Means Segmentation of Breast Parenchymal Tissue in Digital Mammography

Brad Keller^{1,*}, Diane Nathan¹, Yan Wang²,
Yuanjie Zheng¹, James Gee¹, Emily Conant¹, and Despina Kontos¹

¹ Department of Radiology, University of Pennsylvania, Philadelphia, PA, USA
Brad.Keller@uphs.upenn.edu

² Applied Mathematics and Computational Science,
University of Pennsylvania, Philadelphia, USA

Abstract. The relative fibroglandular tissue content in the breast, commonly referred to as breast density, has been shown to be the most significant risk factor for breast cancer after age. Currently, the most common approaches to quantify density are based on either semi-automated methods or visual assessment, both of which are highly subjective. This work presents a novel multi-class fuzzy c-means (FCM) algorithm for fully-automated identification and quantification of breast density, optimized for the imaging characteristics of digital mammography. The proposed algorithm involves adaptive FCM clustering based on an optimal number of clusters derived by the tissue properties of the specific mammogram, followed by generation of a final segmentation through cluster agglomeration using linear discriminant analysis. When evaluated on 80 bilateral screening digital mammograms, a strong correlation was observed between algorithm-estimated PD% and radiological ground-truth of $r=0.83$ ($p<0.001$) and an average Jaccard spatial similarity coefficient of 0.62. These results show promise for the clinical application of the algorithm in quantifying breast density in a repeatable manner.

Keywords: breast cancer, mammography, imaging biomarker, percent breast density, breast cancer risk estimation.

1 Introduction

Beginning with the pioneering work of Wolfe [1], multiple studies have established that the relative amount of fibroglandular tissue seen within the breast, often referred to as breast density, is an image-derived biomarker that has been shown to be an independent risk factor for breast cancer, in fact the most significant after age [2]. Currently, the most commonly used methods to assess breast density rely either on visual assessment by radiologists in distinct categories [3] or through interactive, semi-automated image thresholding [4]. Both approaches are highly subjective with known limitations. Categorical methods, such as with the 4-class Breast Imaging Reporting and Data Systems (BIRADS) [3] system illustrated in figure 1, are associated with only moderate overall agreement, with poor concordance with respect

* Corresponding author.

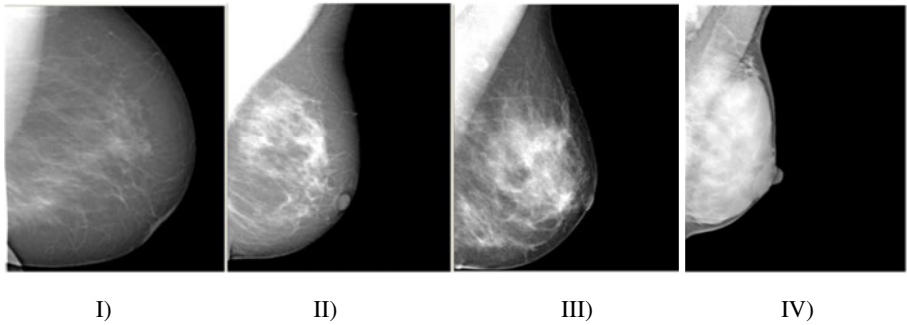


Fig. 1. Sample Digital Mammograms of BIRADS Categories I-IV in Increasing Order of Density. I) $<25\%$ fibroglandular content; II) fibroglandular content between $26-50\%$; III) fibroglandular content between $51-75\%$; IV): fibroglandular content $>75\%$.

to moderately dense breasts versus completely fatty or completely dense breasts [5]. Interactive image-thresholding methods, which require user interaction, are known to introduce reader-variability into density assessment [6].

To address these limitations, automated methods have been proposed for extracting breast density information from digitized film mammography images. For example, Petroudi *et al.* have proposed a density classification scheme based on texture models [7] and Tagliafico *et al.* have investigated adaptive thresholding techniques to identify the fibroglandular tissue regions of a mammogram [8]. However, the translation of these techniques into clinical practice has been limited due to the impracticality of incorporating mammographic film digitization into the clinical workflow solely for the purposes of estimating breast density.

As film mammography is rapidly being replaced by digital mammography, the opportunity arises to develop sophisticated fully-automated algorithms by quantifying breast density directly from the digital images. Digital mammograms capture richer gray-level intensity profiles compared to digitized film images, in which the digitization process also leads to different signal to noise ratio characteristics due to the inherent granularity of the film [9]. Here we propose a novel adaptive multi-cluster fuzzy c-means (FCM) segmentation algorithm for quantifying breast percent density (PD %) from digital mammography images. Our algorithm involves a series of steps; *i*) breast region and pectoral muscle segmentation; *ii*) adaptive histogram-based determination of the optimal number clusters for FCM segmentation; and *iii*) dense tissue cluster merging through a linear discriminant analysis (LDA) agglomeration classifier. The innovation of our algorithm lays in the adaptive nature of the FCM clustering segmentation, which determines the optimal number of clusters based on the breast tissue properties of the specific image, and the agglomeration classifier which combines imaging and patient characteristics to achieve optimal segmentation through cluster merging. We validate our algorithm by comparing to radiologist-provided ground truth of dense tissue on a set of 80 cases with bilateral digital screening mammograms (a total of 160 images) covering the full spectrum of breast densities seen in clinical practice. We compare our algorithm to the standard two-class FCM approach previously used for BIRADS density classification in digitized film mammograms [10].

2 Methods

As a pre-processing step in the proposed algorithm a mask of the breast region is generated, here denoted as M_B , using a previously validated algorithm [11], based on a combination of automated thresholding to identify the breast tissue versus air followed by identification of the pectoral muscle by the straight-line Hough transform (Figures 2a-b). To account for variable gray-level intensities between different mammographic images, the gray level values of the breast region are then normalized using z-score to a (μ, σ) of $(0, 1)$.

2.1 Adaptive Fuzzy C-Means Clustering of Breast Fibroglandular Tissue

Once the breast area is identified, we perform an adaptive k -class fuzzy c-means clustering of the breast region gray-level intensities, where k is optimized for the given image based on the morphology of the normalized histogram of the corresponding image intensity-values. To determine the appropriate k , the grey-level histogram is convolved using a Gaussian kernel of an empirically determined window-size and standard deviation in order to both smooth out quantization noise while simultaneously enhancing concentrations of intensity values. The first and second derivatives of the smoothed histogram are then calculated, and the number of modes (defined as zero-crossings in the first derivative with negative second derivatives) are used to define k ,

$$k = |\{H'(g_n) = 0 : g_n \in M_B; H''(g_n) < 0\}| \quad (1)$$

where $H(g_n)$ is the histogram of the normalized gray-levels values, g_n , for those pixels found in the breast mask, M_B . Once the appropriate number of clusters, k , is determined for a given image of the breast, we then perform k -class FCM clustering [12] on the gray-level values present in the breast region. FCM iteratively optimizes a weighted sum of square error function, which ultimately yields cluster centroids and a cluster-membership matrix for every intensity value in the breast mask. Once this is accomplished, every pixel is assigned to the cluster for which that pixel's intensity value has the highest membership score. An example histogram for a $k=7$ case and the resultant FCM clustering can be seen in Figures 2b and 2c.

2.2 Cluster Agglomeration and Percent Density Calculation

To agglomerate the k -cluster output into a two-class segmentation of the dense versus fatty tissue, an LDA classifier is trained to determine the cluster cutoff for dense tissue. All clusters described by centroids of equal or higher intensity than the LDA cutoff are then combined (i.e., agglomerated) into a single segmented dense tissue region, M_D .

Predictor variables of the LDA-classifier included image histogram statistics, which are often used to classify images into BIRADs categories [10]; image acquisition parameters and patient characteristics, which have been shown to correlate to PD% [13]; and parameters of the output of the FCM clustering. In order to reduce

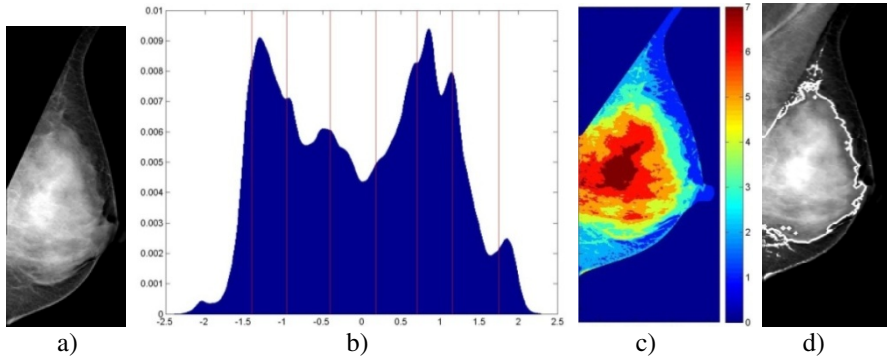


Fig. 2. Segmentation algorithm stages for a $k=7$ mammogram. a) Segmented breast region; b) Normalized breast-pixel intensity histogram with FCM cluster centroids (vertical lines); c) Pixel cluster-membership represented by shading; d) Final dense tissue segmentation.

the dimensionality of the input predictor variables, stepwise feature selection is performed, where terms are systematically added to the LDA-model classifier based on the relative explanatory power of the terms as described by Draper *et al.* [14]. The final feature set included: the third central moment and 5th percentile of the histogram, kurtosis, mean and standard deviation of the un-normalized histogram, age, breast thickness, mammogram paddle compression-force, KVP, exposure, and parameters of a sigmoid-curve fit to horizontal and vertical extents of the resulting dense-tissue segmentation as a function increasing the cut-off cluster choice. An example of the agglomeration result can be seen in Figure 2d. From this final density mask, we calculate mammographic percent density, PD%, by computing

$$PD\% = \frac{|M_D|}{|M_B|} \cdot 100\% \quad (2)$$

2.3 Dataset and Algorithm Evaluation

To validate our proposed algorithm, we identified 80 cases with bilateral MLO-view post-processed digital mammography images (*PremiumView*TM, GE Healthcare), which yielded a total of 160 images for analysis. All images were acquired using a standard screening protocol on a Senograph DS (GE Healthcare) full-field digital mammography (FFDM) system, with an isotropic 100 μ m resolution. For each image, a trained breast imaging radiologist provided a breast PD% estimation and dense tissue segmentation using a validated user-interactive image-thresholding tool for breast PD% estimation (*Cumulus*, Ver. 4.0, Univ. Toronto) [15].

For our experiments, the agglomeration LDA-classifier, including the feature selection stage, is trained using a leave-one-woman-out (2 mammograms) schema on the output of the unsupervised FCM clustering to select the optimal cutoff point for dichotomizing the image into fatty and dense segmentations, in this case the lowest intensity cluster that is still a “dense” cluster, in order to maximize the agreement between computer-estimated and radiologist-defined PD%. To evaluate the accuracy

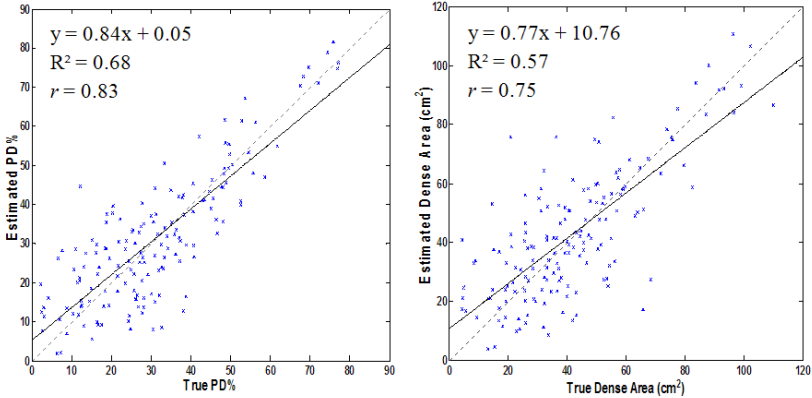


Fig. 3. Scatter-plots of Algorithm-Estimated vs. Radiologist-Provided PD% (Left) and Dense-tissue Segmentation Area (Right). Regression-equations, R^2 , Pearson Correlations, the linear regression line (black) and the identity-lines (dashed-gray) are provided for reference.

of our algorithm in estimating breast PD%, we compute the Pearson product-moment correlation coefficient [16], r , between the algorithm-estimated PD% and the radiologist-provided PD%, considered here as ground-truth for our experiments. Spatial agreement between the algorithm-segmented and radiologist-segmented dense tissue regions is evaluated using the Jaccard index, J , [17] defined as

$$J = \frac{R_D \cap M_D}{R_D \cup M_D} \tag{3}$$

where R_D and M_D are the fibroglandular tissue segmentations generated by the radiologist and algorithm respectively. Finally, we also compare our algorithm with the standard two-class FCM segmentation that has been previously used for dense tissue segmentation in digitized film mammograms [10].

3 Results

For our experiments, image intensity histograms were constructed with a fine bin-width of 0.01 on the z-score normalized intensity histogram. With regards to cluster-count optimization, it was found that distribution of frequencies of k was approximately Gaussian, centered at $k=6$. We applied a bounding constraint on k to be between 2 (as there will always be an adipose and a fibroglandular tissue cluster) and 9 (for speed and memory considerations). Computation of k was not found to be particularly sensitive to changes in histogram construction or small peaks.

Feature selection during LDA-training was found to select the exact same feature set for all 80 iterations of the leave-one-woman-out training, indicating that the features used were robust to case variation and provide orthogonal information. When comparing radiologist-derived PD% to algorithm-estimated PD%, we were able to obtain a Pearson correlation of $r=0.83$. Correlation between estimated and true breast

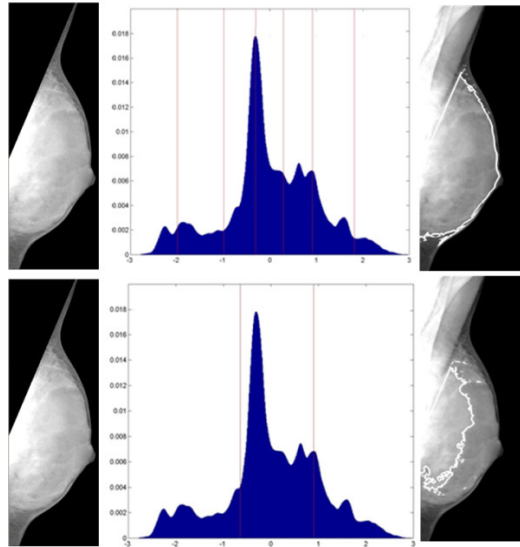


Fig. 4. Comparison between 2-class (top) and adaptive $k=6$ -class (bottom) FCM segmentation of a BIRADS-IV category breast. Left) Breast Mask; Center) Gray-level Histogram with marked FCM-centroids; Right) Final segmentation

area calculations was found to be $r=0.99$, while the correlation between estimated and true absolute dense-tissue area was found to be $r=0.75$. Scatter-plots of the estimated vs. true PD% and dense area are provided in figure 3. Good spatial agreement between true and algorithm-derived dense area also found, with a corresponding distribution of Jaccard indices of $J=0.62\pm 0.22$ for the 160 density segmentations. The 2-class FCM, previously used for segmenting dense tissue in *film* mammography, showed a low correlation of $r=0.05$ ($p>0.1$) between 2-class FCM estimated PD% and ground truth as well as lesser spatial agreement of $J=0.57\pm 0.24$.

4 Discussion

The proposed fully-automated algorithm was successful in identifying the fibroglandular tissue of the breast in digital mammographic images. Strong agreement with radiologist-provided ground truth was obtained, both in terms of the quantitative PD% estimate and the spatial agreement of the segmented dense tissue. Furthermore, the spatial agreement between automated-algorithm segmentations and radiologic ground-truth ($J=0.62\pm 0.22$) was found to be similar to the human-observer inter-reader variability of 0.65 ± 0.18 previously reported by Bakic *et al.* [15]. These findings indicate that the proposed algorithm can provide clinically relevant information from digital mammography for the assessment of breast density.

One surprising result was the relatively poor performance of the two-class fuzzy-c-means paradigm, previously used for density assessment in digitized film mammograms, in identifying the dense tissue region. Although BIRADS classification and

not PD% quantification has often been the primary focus [10], this finding is in apparent contrast to findings by Torrent et al. who reported good visual agreement between FCM and expert markings [18]. Further investigation showed that one possible explanation for the poor performance of the 2-class FCM algorithm is the fact that majority gray-level intensity profiles of breast tissue as extracted from digital mammograms tend to be multi-modal, such as in the BIRADS-IV case illustrated in Figure 4. Given that appropriate selection of the number of clusters, k , is critical for proper clustering, the intensity profile of digital mammograms are complex enough that multi-class techniques, such as the one described in this work, may be required to appropriately analyze digital mammography images.

As breast tissue seen mammographically is a 2D superimposition of different tissue types with different image properties, future studies should seek to expand density-based risk-stratification analysis beyond the dichotomous, fatty vs. dense tissue paradigm. Furthermore, as mammography is essentially limited by the effect of projection/tissue superimposition, volumetric analysis of fibroglandular tissue through emerging tomographic breast imaging modalities, such as breast tomosynthesis and magnetic resonance imaging, has been suggested as necessary to advance breast-cancer risk modeling [19] and the approach described in this work could become the foundation for fully-automated density quantification from three-dimensional images.

5 Conclusion

We have proposed and demonstrated the efficacy of a novel fully automated algorithm for fibroglandular tissue segmentation in digital mammography. We were able to obtain strong correlation between the output of the computerized algorithm and radiologist-provided ground truth. These results show promise for the potential clinical relevance and applicability of our method to quantify breast density in a repeatable and objective manner. This fully-automated method could accelerate the clinical translation of density-based cancer risk stratification and pave the way for new personalized screening and prevention strategies.

Acknowledgements. This work was supported by ACS Grant RSGHP-CPHPS-119586 and DOD Grant BC086591. The authors would also like to thank Dr. Martin Yaffe (University of Toronto) for providing *Cumulus* software tools as well as for useful discussions.

References

1. Wolfe, J.N.: Breast patterns as an index of risk for developing breast cancer. *Am. J. Roentgenol.* 126, 1130–1137 (1976)
2. Tice, T.A., Kerlikowske, K.: Screening and prevention of breast cancer in primary care. *Prim. Care* 36, 533–558 (2009)
3. D’Orsi, C.J., Bassett, L.W., Berg, W.A., Feig, S.A., Jackson, V.P., Kopans, D.B.: Breast imaging reporting and data system: ACR BI-RADS-mammography, 4th edn. Reston. Am. Col. of Rad. (2003)

4. Martin, K.E., Helvie, M.A., Zhou, C., Roubidoux, M.A., Bailey, J.E., Paramagul, C., Blane, C.E., Klein, K.A., Sonnad, S.S., Chan, H.P.: Mammographic density measured with quantitative computer-aided method: comparison with radiologists' estimates and BI-RADS categories. *Radiology* 240, 656–665 (2006)
5. Nicholson, B.T., LoRusso, A.P., Smolkin, M., Bovbjerg, V.E., Petroni, G.R., Harvey, J.A.: Accuracy of assigned BI-RADS breast density category definitions. *Acad. Radiol.* 13, 1143–1149 (2006)
6. Yaffe, M.J.: Mammographic density. Measurement of mammographic density. *Breast Cancer Res. Treat.* 10, 209–218 (2008)
7. Petroudi, S., Kadir, T., Brady, M.: Automatic classification of mammographic parenchymal patterns: A statistical approach. Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2003, vol.1, pp. 798–801 (2003)
8. Tagliafico, A., Tagliafico, G., Tosto, S., Chiesa, F., Martinoli, C., Derchi, L.E., Calabrese, M.: Mammographic density estimation: comparison among BI-RADS categories, a semi-automated software and a fully automated one. *Breast* 18, 35–40 (2009)
9. Yaffe, M.: Digital mammography. In: PACS, pp. 363–371. Springer, Heidelberg (2006)
10. Oliver, A., Freixenet, J., Martí, R., Pont, J., Pérez, E., Denton, E.R.E., Zwiggelaar, R.: A novel breast tissue density classification methodology. *IEEE Trans. Inf. Technol. Biomed.* 12, 55–65 (2008)
11. Karssemeijer, N.: Automated classification of parenchymal patterns in mammograms. *Phys. Med. Biol.* 43, 365–378 (1998)
12. Bezdek, J.: Pattern recognition with fuzzy objective function algorithms. Kluwer Academic Publishers, Norwell (1981)
13. Lu, L.J.W., Nishino, T.K., Khamapirad, T., Grady, J.J., Leonard, M.H., Brunder, D.G.: Computing mammographic density from a multiple regression model constructed with image-acquisition parameters from a full-field digital mammographic unit. *Phys. Med. Biol.* 52, 4905–4921 (2007)
14. Draper, N.R., Smith, H.: Applied Regression Analysis. Wiley-Interscience, Hoboken (1998)
15. Bakic, P.R., Carton, A.K., Kontos, D., Zhang, C., Troxel, A.B., Maidment, A.D.A.: Breast percent density: estimation on digital mammograms and central tomosynthesis projections. *Radiology* 252, 40–49 (2009)
16. Pearson, K.: Mathematical contributions to the theory of evolution. III. Regression, heredity and panmixia. *Philos. Trans. Royal Soc. London Ser. A* 187, 253–318 (1896)
17. Jaccard, P.: The Distribution of Flora in the Alpine Zone. *New Phytologist* 11, 37–50 (1912)
18. Torrent, A., Bardera, A., Oliver, A., Freixenet, J., Boada, I., Feixes, M., Martí, R., Lladó, X., Pont, J., Pérez, E., Pedraza, S., Martí, J.: Breast Density Segmentation: A Comparison of Clustering and Region Based Techniques. In: Krupinski, E.A. (ed.) *IWDM 2008*. LNCS, vol. 5116, pp. 9–16. Springer, Heidelberg (2008)
19. Kopans, D.B.: Basic physics and doubts about relationship between mammographically determined tissue density and breast cancer risk. *Radiology* 246, 348–353 (2008)

Learning Image Context for Segmentation of Prostate in CT-Guided Radiotherapy

Wei Li^{1,2}, Shu Liao¹, Qianjin Feng², Wufan Chen², and Dinggang Shen^{1,*}

¹ IDEA Lab, Department of Radiology and BRIC,
University of North Carolina at Chapel Hill, USA

² Biomedical Engineering College, Southern Medical University, Guangzhou, China
dgshen@med.unc.edu

Abstract. Segmentation of prostate is highly important in the external beam radiotherapy of prostate cancer. However, it is challenging to localize prostate in the CT images due to low image contrast, prostate motion, and both intensity and shape changes of bladder and rectum around the prostate. In this paper, an online learning and patient-specific classification method based on location-adaptive image context is proposed to precisely segment prostate in the CT image. Specifically, two sets of position-adaptive classifiers are respectively placed along the two coordinate directions, and further trained with the previous segmented treatment images to jointly perform the prostate segmentation. In particular, each location-adaptive classifier is recursively trained with different image context collected at different scales and orientations for better identification of each prostate region. The proposed learning-based prostate segmentation method has been extensively evaluated on a large set of patients, achieving very promising results.

Keywords: Radiotherapy, Prostate segmentation, Classification, Image context.

1 Introduction

Prostate cancer is one of the most common cancers in males and is a leading cause of male cancer death in US [1]. The external beam radiation treatment can provide a non-invasive and effective therapy for prostate cancer. It is usually planned on a planning CT image, namely *planning image*, on which the prostate and nearby critical structures are manually contoured. Then, the treatment is delivered in daily fractions during a period of eight to nine weeks. At each treatment day, a new CT image, namely *treatment image*, is acquired to guide the dose delivery, with goal of maximizing the dose delivered to the prostate and minimizing the dose delivered to healthy tissue such as bladder and rectum. Therefore, the success of image-guided radiotherapy (IGRT) highly depends on the accurate localization of prostate.

However, the prostate segmentation from CT images is challenging mainly due to three factors. *First*, prostate has low intensity contrast in the CT images. *Second*, the prostate motion/deformation (mainly caused by both water in the bladder and air in

* Corresponding author.

the rectum) is complex and unpredictable, even after rigid alignment of treatment images to the planning image based on the bone structures. *Third*, the uncertain presence of rectum gas makes the appearance of treatment images inconsistent, as demonstrated in Fig.1.

So far, several categories of methods have been proposed for prostate segmentation from CT images. The first category of segmentation methods is the deformable-model based segmentation [2-4]. The second category of segmentation methods is the statistical-atlas based segmentation with registration techniques [5, 6]. Also, other methods have been proposed based on different optimization strategies [7, 8], or working on different imaging modalities, e.g., MR and US [9]. In this paper, we focus on developing a learning-based method to segment prostate from CT images.

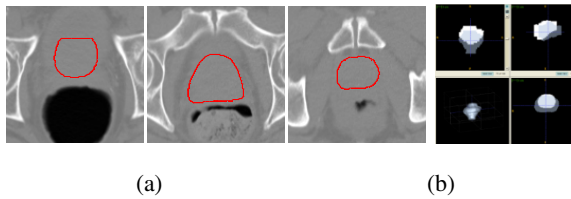


Fig. 1. (a) Slices showing the prostate and surrounding organs (i.e., bladder, rectum). The black parts are the air in the rectum, and the red curves are the boundaries of prostate manually delineated by a radiologist. (b) 3D-display showing large motion/deformation of prostate in the two treatment images of the same patient, even after bone-based rigid alignment. Here, white and grey represent the two prostates after bone-based registration.

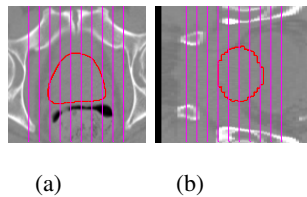


Fig. 2. Illustration of location-adaptive classifiers placed along the two coordinate directions around prostate region. Each red straight line represents the place where we will place a location-adaptive classifier. The red curves represent the boundaries of prostate in (a) axial and (b) sagittal slices.

It is well known that the context and high-level information plays an important role in object recognition and scene understanding, especially for medical image analysis [10, 11]. Recently, a supervised machine learning method with context information has been proposed in [12]. In this work, the discriminative probability maps created by the learned classifier are used as context information recursively. Motivated by [12], we propose a patient-specific classification method for segmentation of prostate using context and appearance features under the online-learning framework. In particular, we improve the segmentation of prostate by learning both context and appearance information from the previous treatment images, thus achieving accurate segmentation for new treatment images of the same patient.

Moreover, we further propose a location-adaptive classification method to improve the classification accuracy of prostate. This is motivated by the fact that a single learned global classifier and its model parameters (trained globally through the entire volume) may not properly reflect the features of the local regions. Some similar ideas of employing distributed local classifiers have been used for medical image segmentation [13]. Specifically, we will train two sets of location-adaptive classifiers along the two coordinate directions for a 3D region of interest (ROI) around prostate, as shown in Fig. 2 (a) and (b). Each location-adaptive classifier is trained only for some 2D slices within in the ROI. The classification results with the two sets of location-adaptive classifiers are fused into a final 3D classification result and further segmented by a level-set based method.

2 Method

Our approach consists of an online training stage and a classification stage in the course of radiotherapy for the patient. In the training stage, two sets of location-adaptive classifiers are initially trained by the patient's first 3 daily scans with manual segmentations, which can be further online-updated by the latest acquired scans (e.g., 4 scans) of the same patient during the radiotherapy. With more and more subsequent treatment images are acquired and segmented, our classifiers can be updated with the automatic segmentation results of the latest-acquired treatment images (e.g., 4 images), along with the first 3 daily scans with manual segmentations. In this way, our trained classifiers can track the up-to-date patient-specific image context and appearance to improve the classification of prostate in the next daily treatment image. Note that our approach can work fully automatically after the initialization.

Specifically, to allow our classifiers to learn the image context and appearance in the training images, besides normalizing their intensities, the patient's pose in these images need to be aligned. We can use the pelvic bones extracted (with simple thresholding) from the training images to rigidly align all the training images onto the planning image space. For saving the time for both the training and testing stages, we further extract a ROI around the prostate, i.e., shown in Fig. 2. Then, to train each location-adaptive classifier, we just collect image features from the corresponding slices that this classifier is responsible for, and then use a machine learning technique [12] to build this classifier, along with their respective segmentation labels. Finally, these trained classifiers can be used to provide a probability segmentation map for prostate in the new treatment image, which can be further transformed into a binary segmentation by a level-set based technique. Note that both the treatment images and their respective segmentation results by our method can be used to update the training dataset for online refinement of our classifiers.

In the segmentation stage, each new treatment image is first aligned onto the planning image space. Then, the ROI used in the training stage will be used to extract the target region in the current aligned treatment image, which will then be classified/labeled by the two sets of our trained location-adaptive classifiers. The two prostate probability maps produced by the two sets of classifiers can be fused into a single probability map and further transformed into a binary segmentation by a level-set based technique [14]. Finally, the binary prostate segmentation result, along with its corresponding intensity image, can be used to update the classifiers.

2.1 Auto-context Classification Formulation

In this section, we briefly present the formulation of auto-context classification and notations. More details can be referred to [12]. We view an image as a function X from the spatial domain $\dot{U} \subset \mathbb{R}^2$ to an intensity value in \mathbb{R} . For each image X , its corresponding ground-truth segmentation/label image Y is also available (i.e., manual prostate segmentation), with value of each pixel belonging to $\{0,1\}$. There are M images used for training. And the j -th pixel of image X_i ($i = 1 \cdots M$) is represented by x_{ij} , where $j \in \{1 \cdots N\}$ and N is the number of pixels in image X_i . Then, the training set can be represented by $S = \{(x_{ij}, y_{ij}), i = 1 \cdots M, j = 1 \cdots N\}$, where y_{ij} is the corresponding label in the label image Y_i .

We aim to learn a classifier to find an optimal classification result through Maximum a Posteriori (MAP) given a new image X , i.e. $Y^* = \operatorname{argmax}(p(Y|X))$. Here, the marginal distribution is employed to solve $p(Y|X)$. To better approximate the marginal distribution, an auto-context model is proposed, in which the classification map produced by a traditional classifier, such as SVM or AdaBoost, is integrated into the training data set recursively to input context information. The iterative algorithm updates $p^{(t+1)}(y_j|B_j(X), C^t)$ to approach $p(y_j|X)$ asymptotically, where C^t is the context features at the t -th iteration and $B_j(X)$ denotes all pixels in an image patch centered at the j -th pixel.

2.2 Location-Adaptive Classifier

The identification of prostate, bone, air, and other tissue regions is important for training the location-adaptive classifiers. To achieve this, we use the segmentation labels of training images to locate the prostate, and identify (i.e., with simple thresholding) bone, air and other tissue regions on the training images. After that, we randomly sample pixels from each of the above regions as training instances. Then, the image appearance features of these training instances are computed and denoted as A_{kl} ($k = 1 \cdots K, l = 1 \cdots L$), where K is the number of slices used for training a specific location-adaptive classifier, and L is the number of training instances. Context features are computed from the classification map by using a pattern that includes a large number of locations, i.e., shown in Fig. 3(b). These context features are denoted by C_{kl}^t ($k = 1 \cdots K, l = 1 \cdots L, t = 0 \cdots T$). Finally, we can iteratively train our location-adaptive classifier. At the end of the training procedure, we get a sequence of classifiers L_t ($t = 0 \cdots T$) for a specific location-adaptive classifier.

2.3 Feature Extraction

Two types of features are employed in our learning-based segmentation algorithm. The first one is the appearance feature, calculated from the original image. The second one is the context feature, computed from the classification map. The appearance features include a set of rectangular Haar features as used in [15]. Haar features are widely used in object recognition, which can be computed at different scales and with high speed by using integral images. We extract Haar features from different scales to facilitate our prostate classification. Fig. 3(a) shows some 2D Haar-like features used in our algorithm. Other appearance features adopted are the

histogram of oriented gradient (HOG) features[16]. The patch size for computing the above appearance features is 21×21 . Finally, the coordinate of each pixel is also included as feature for learning the movement information of prostate in the pelvic.

Context features are updated recursively based on the classification maps which are produced iteratively. For each pixel of interest, a number of rays in equal-degree intervals are extended out from the current pixel and we sparsely sample the context locations on these rays, as shown in Fig. 3(b). For each context location, their classification probabilities and the mean probability within 3×3 windows are included as context features. We consider that the context information is also from the intensity variation in these selected context locations. Thus, the absolute intensity variation between the context location and current pixel of interest is also included as context features in our algorithm. Then, the classifiers can learn location information from the air or bone regions (which often have large intensity variation from the current pixel of interest) through our context features. These context features implicitly represent the shape and configuration information of the prostate.

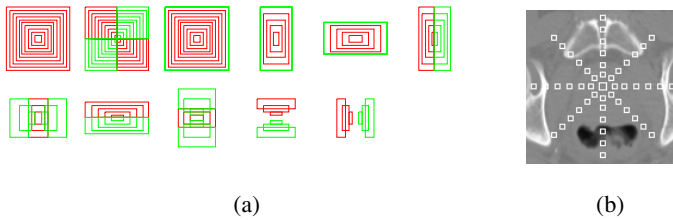


Fig. 3. (a) Examples of some 2D Haar-like features. (b) Context sample pattern

3 Experiments

Our data consists of 11 patients, each with more than 9 daily CT scans, with total image count of 161. Each image has resolution of $1 \times 1 \times 3 \text{ mm}^3$. The first 3 images of each patient are used to initialize the patient-specific learning process. The expert manual segmentation results are also available for each image. Since the classifiers are designed to learn the patient-specific information, all training and testing experiments are done on the images from the same patient. Hereafter, we index all images starting from number 1, i.e., the planning image will be image 1 and the first treatment image will be image 2, and so on. Three quantitative measures are used to evaluate the performance of our algorithm by comparing the automated segmentation results with the manual segmentations, and they are: Dice similarity coefficient (DICE), average surface distance (ASD), and centroid distance (CD).

Fusion of results along different coordinate directions. To validate the effectiveness of the fusion on the two results obtained from two classifier sets along two coordinate directions, we carried out a visual comparison between the results obtained from classifiers along each direction and the results obtained by the fusion. The fusion result was obtained by averaging the two classification results; then, a level-set based technique was used to extract smooth prostate boundary surface from the fused result.

Fig. 4 shows a sample fusion result. From Fig. 4, it can be observed that the fusion result is better than the results obtained along two directions separately.

Evaluation of performance on different number of training images. Here, we tested the performance of our proposed method under different number of training images. As examples, we selected 3 patients (patient ID 3, 5 and 10) with larger inter-treatment segmentation performance, e.g., these three patients have large variations of Dice measures compared to other patients (see Fig. 6(a)). Next, we selected one particular image from each of these three selected patients, i.e., image 11 of patient 3, image 12 of patient 5, and image 14 of patient 10. Then, we tested on each of these 3 images by using our location-adaptive classifiers trained by the latest 2 to 6 treatment images of the same patient. The Dice measures are shown in Fig. 6(c). It can be observed that the Dice measures become steady when the number of training images is 4 or larger. Therefore, the number of the training images is set to 3~7 for the tradeoff between training time and segmentation accuracy in all our experiments.

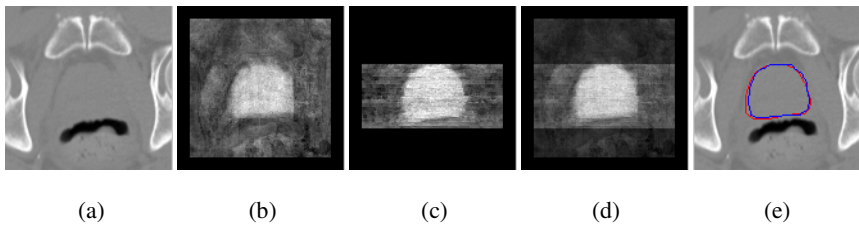


Fig. 4. Demonstration on the fusion of classification results along two coordinate directions. (a) Original image; (b) Classification result along z coordinate; (c) Classification result along y coordinate; (d) Fusion of (b) and (c); (e) Ground truth (red) and our estimated result (blue).

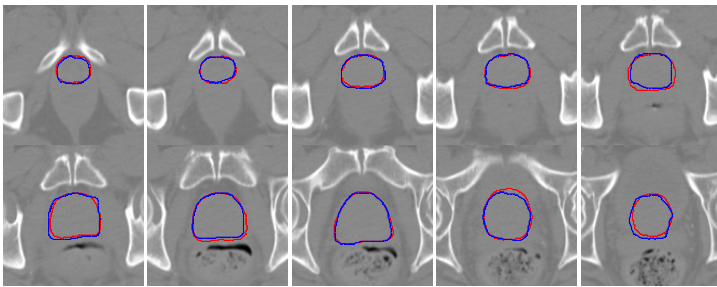


Fig. 5. Segmentation results on selected slices of image 11 of patient 5 (Dice 0.906). The red contours show the manually delineated boundaries of prostate by a radiation oncologist, while the blue contours show the results of the proposed method.

Evaluation of online learning mechanism. Since our method is learning-based, the number of training data cannot be too small. Thus, we use the first 3 images with manual segmentation and the latest 4 images with automatic segmentations to train our classifiers. In this way, our location-adaptive classifiers can learn the up-to-date information and reserve the information from manual segmentations of radiologist as

well. Fig. 5 shows an example on segmentation results for a patient image. The results on all images of all 11 patients are shown in Fig. 6(a) for Dice measures and Fig. 6(b) for ASD measures. Also, the centroid differences in lateral (x), anterior-posterior (y), and superior-inferior (z) directions are shown in Fig. 7. Three existing state-of-the-art prostate segmentation algorithms [3], [5] and [2] are compared with our method. Our mean Dice measure and mean ASD on all patients are 0.908 and 1.40 mm, respectively, which are better than the mean Dice measure of 0.893 and the mean ASD of 2.08 mm reported in [3]. Our mean centroid differences are 0.18 mm, -0.02 mm and 0.57 mm along x, y and z directions, which are comparable to the best result of -0.26 mm, 0.35 mm, and 0.22 mm reported in [5], while our mean Dice measure 0.908 is significantly higher than theirs 0.82. Our median probability of detection and false alarm are 0.90 and 0.10, respectively, which are much better than 0.84 and 0.13 reported in [2].

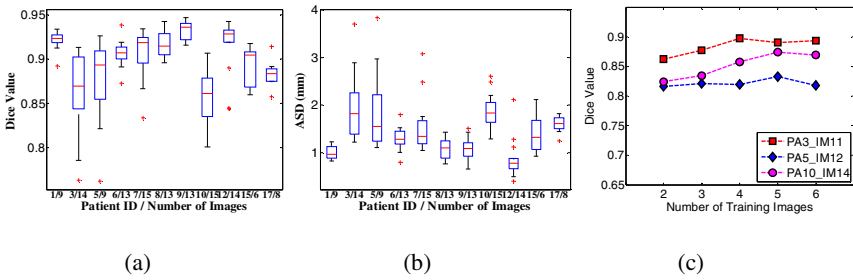


Fig. 6. (a) Dice measure of results and (b) average surface distance (ASD). (c) Change of the performance of our proposed method w.r.t. the different number of images used for training. Here, “PA3_IM11” means the image 11 of patient 3.

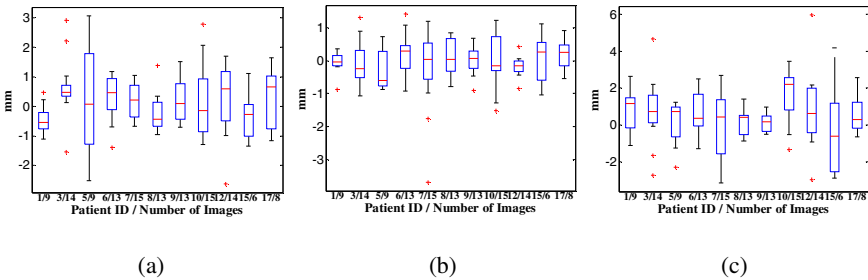


Fig. 7. (a), (b) and (c) are the centroid differences in lateral (x), anterior-posterior (y), and superior-inferior (z) directions, respectively. The symbols in the figure are the same as Fig. 6.

4 Conclusion

We have proposed a patient-specific online-learning classification method based on image context information, for segmenting prostate in the CT scans during radiotherapy. Our method uses only the training samples from the previous treatment

images of the same patient to specifically learn the patient-specific information for enhancing the prostate segmentation result, which is different from the conventional deformable-model or image-registration based methods that generally use the population information to guide the segmentation. Our trained classifiers can be updated online by the latest-acquired treatment images, thus better adapting to the prostate change of the current patient during the treatment. Experimental results show that our proposed method can produce prostate segmentations more accurately from clinical CT images, than the other state-of-the-art segmentation algorithms.

Acknowledgments. This work was supported by NIH grant R01 CA140413 and the National Basic Research Program of China (973 Program) (No. 2010CB732505). The image data and expert contours used in this study were provided Dr. E.L. Chaney in the Department of Radiation Oncology, UNC-Chapel Hill.

References

1. American Cancer Society.: Cancer facts and figures 2006. In: American Cancer Society (ed.). American Cancer Society, Atlanta (2006)
2. Chen, S., Lovelock, D.M., Radke, R.J.: Segmenting the prostate and rectum in CT imagery using anatomical constraints. *Med. Image Anal.* 15, 1–11 (2011)
3. Feng, Q., Foskey, M., Chen, W., Shen, D.: Segmenting CT prostate images using population and patient-specific statistics for radiotherapy. *Med. Phys.* 37, 4121–4132 (2010)
4. Freedman, D., Radke, R.J., Zhang, T., Jeong, Y., Lovelock, D.M., Chen, G.T.: Model-based segmentation of medical imagery by matching distributions. *IEEE Trans. Med. Imag.* 24, 281–292 (2005)
5. Davis, B.C., Foskey, M., Rosenman, J., Goyal, L., Chang, S., Joshi, S.: Automatic segmentation of intra-treatment CT images for adaptive radiation therapy of the prostate. *Med. Image Comput. Comput. Assist. Interv.* 8, 442–450 (2005)
6. Wang, H., Dong, L., Lii, M.F., Lee, A.L., de Crevoisier, R., Mohan, R., Cox, J.D., Kuban, D.A., Cheung, R.: Implementation and validation of a three-dimensional deformable registration algorithm for targeted prostate cancer radiotherapy. *Int. J. Radiat. Oncol. Biol. Phys.* 61, 725–735 (2005)
7. Ghosh, P., Mitchell, M.: Segmentation of medical images using a genetic algorithm. In: Proceedings of the 8th Annual Conference on Genetic and Evolutionary Computation, pp. 1171–1178. ACM, Seattle (2006)
8. Song, Q., Wu, X., Liu, Y., Smith, M., Buatti, J., Sonka, M.: Optimal graph search segmentation using arc-weighted graph for simultaneous surface detection of bladder and prostate. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5762, pp. 827–835. Springer, Heidelberg (2009)
9. Zhan, Y., Shen, D.: Deformable segmentation of 3-D ultrasound prostate images using statistical texture matching method. *IEEE Trans. Med. Imag.* 25, 256–272 (2006)
10. Hupse, R., Karssemeijer, N.: Use of normal tissue context in computer-aided detection of masses in mammograms. *IEEE Trans. Med. Imaging* 28, 2033–2041 (2009)
11. Oliva, A., Torralba, A.: The role of context in object recognition. *Trends Cogn. Sci.* 11, 520–527 (2007)

12. Tu, Z., Bai, X.: Auto-Context and Its Application to High-Level Vision Tasks and 3D Brain Image Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 32, 1744–1757 (2010)
13. Richard, N., Dojat, M., Garbay, C.: Distributed Markovian segmentation: Application to MR brain scans. *Pattern Recogn.* 40, 3467–3480 (2007)
14. Chan, T.F., Vese, L.A.: Active contours without edges. *IEEE Transactions on Image Processing* 10, 266–277 (2001)
15. Viola, P., Jones, M.J.: Robust Real-Time Face Detection. *Int. J. Comput. Vision* 57, 137–154 (2004)
16. Dalal, N., Triggs, B.: Histograms of Oriented Gradients for Human Detection. In: *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005)*, vol. 1, pp. 886–893. IEEE Computer Society, Los Alamitos (2005)

Joint Thrombus and Vessel Segmentation Using Dynamic Texture Likelihoods and Shape Prior

Nicolas Brieu¹, Martin Groher¹, Jovana Serbanovic-Canic^{2,3},
Ana Cvejic^{2,3}, Willem Ouwehand^{2,3,4}, and Nassir Navab¹

¹ Computer Aided Medical Procedures, Technische Universität München, Germany
brieu@in.tum.de

² The Wellcome Trust Sanger Institute, Hinxton, UK

³ Department of Hematology, University of Cambridge, UK

⁴ NHS Blood and Transplant, Cambridge, UK

Abstract. The segmentation of thrombus and vessel in microscopic image sequences is of high interest for identifying genes linked to cardiovascular diseases. This task is however challenging because of the low contrast and the highly dynamic conditions observed in time-lapse DIC *in-vivo* microscopic scenes. In this work, we introduce a probabilistic framework for the joint segmentation of thrombus and vessel regions. Modeling the scene with dynamic textures, we derive two likelihood functions to account for both spatial and temporal discrepancies of the motion patterns. A tubular shape prior is moreover introduced to constrain the aortic region. Extensive experiments on microscopic sequences quantitatively show the good performance of our approach.

1 Introduction

Thrombosis refers to the aggregation of blood cells occurring after a blood vessel injury. The analysis of thrombus formation in mutant zebrafish larvae is becoming crucial to identify genes related to cardiovascular diseases [11]. The following experiment is conducted: the specimen is put under a microscope, the wall of its caudal aorta is artificially injured to initiate thrombosis, and images are recorded using a microscope/camera system. Measures such as time to attachment of the first blood cell, thrombus surface area, and time to occlusion of the aorta, are manually computed to characterize the thrombus formation. Since such experiments are repeated many times to conduct statistical studies, the characterization needs to be automated. This can be achieved by developing algorithms for the segmentation of the thrombus and of the aortic regions.

Segmenting *in vivo* microscopic images is challenging. As illustrated by previous works on microscopic image segmentation [6,9,13], common intensity-based or gradient-based methods can hardly be applied because of the low image quality and the cluttered environment of *in vivo* scenes. This is particularly true in our application (cf. Fig. 1a). The low contrast between the thrombus and the aorta challenges the computation of intensity-related features. The number of blood cell or streamline-related clutters in the aorta hampers the derivation of

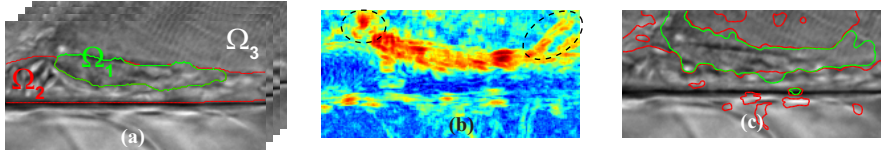


Fig. 1. (a) Microscopic image sequence and Ground-truth segmentation; (b) Residual energy from DT modeling [4]; (c) Intensity-based segmentation of (b)

gradient-related cues. As such, motion seems to be a better discriminative cue. To account for the non-homogeneity of biological objects which hampers the use of first-order intensity-related features, a motion-related flux tensor derived from optical-flow computation has been introduced in [12]. In our case, however, the chaotic movement of the blood cells in the uncoagulated aorta contradicts the brightness constancy assumption made for optical-flow derivation. As such, and similarly to [3], we favor the Dynamic Texture (DT) approach of Doretto et al. [7] that provides reliable motion-related cues in similarly complex scenarios [4, 8] (Fig. 1b-c).

Our previous work [3] has two main limitations. First, it is restricted to the segmentation of the thrombus region. Second, it is disrupted by the slow circulation of blood cells in the collateral vessels and by the local non-rigid movements in the background. In this paper, we propose an approach which overcomes these limitations. We derive likelihood and prior energies from the following observations. (1) The thrombus, the uncoagulated aorta, and the background are characterized by distinct motion patterns. We model the spatial homogeneity of the motion patterns in each region with a *motion-segmentation* likelihood energy. (2) The thrombus is the only region which does not exist at the initial state of the sequence, i.e. before the attachment of the first blood cell. We introduce a second likelihood energy, the *event-detection* energy, which characterizes the variations of the motion patterns between the initial and the current states. (3) The spatial extents of the regions are constrained: the aorta resembles a tubular shape and the thrombus is included in the aorta. By modeling these constraints as a shape prior energy, we counter-balance the collateral vessels and further increase the accuracy of the segmentations.

The key contributions of this paper are the introduction of the *event-detection* likelihood energy, the definition of the shape prior energy, and the quantitative validation of our method on *in vivo* microscopic sequences against manually delineated ground truth segmentations. Experiments show the improvements in accuracy induced by the two newly developed energies.

2 Method

Let us consider a video sequence $\mathcal{S} = \{\mathcal{I}^p, 0 \leq p < T\}$. For each time p , our objective is to partition the image domain Ω into the thrombus (Ω_1), the uncoagulated aorta (Ω_2), and the background region (Ω_3) (Fig. 1a). The reunion

of the thrombus and of the uncoagulated aortic regions forms the aortic region, $\Omega_{12} = \Omega_1 \cup \Omega_2$. To assess the temporal information necessary to define the two likelihood energies, we consider the sub-sequence of τ frames centered at time p , $\mathcal{S}^p = \{\mathcal{I}^t, p - \tau/2 \leq t \leq p + \tau/2\}$, and also the initial sub-sequence \mathcal{S}^0 . Finding the optimal partition ω^* of the image domain at time p given \mathcal{S}^p and \mathcal{S}^0 is formulated as an energy minimization problem

$$\omega^* = \arg \min_{\omega} [E(\mathcal{S}^p, \mathcal{S}^0 | \omega) + E(\omega)]. \tag{1}$$

The likelihood energy $E(\mathcal{S}^p, \mathcal{S}^0 | \omega)$ regroups both the *motion-segmentation* and the *event-detection* energies. The prior energy $E(\omega)$ constrains the partition.

2.1 Dynamic Texture Likelihoods

The definition of motion-based segmentation likelihoods can be decomposed in a series of successive steps. We have to select a motion model, to identify for each pixel the model parameters corresponding to the observed motion pattern, and to define a metric on the model parameter subspace. This permits the relative comparison of the motion patterns observed in the scene and the introduction of likelihood functions accounting for their homogeneity in space and in time.

Dynamic Texture Model. Let us consider the sub-sequence \mathcal{S}^p and the square spatial patch of m pixels centered at the pixel \mathbf{x} . The intensities $\mathbf{y}(t) \in \mathbb{R}^m$ observed in the patch at a time t are related to a set of hidden variables $\mathbf{z}(t) \in \mathbb{R}^n$ by the *linear dynamic system* $\mathbf{y}(t) = \mathbf{Cz}(t) + \mathbf{w}(t)$ and $\mathbf{z}(t+1) = \mathbf{Az}(t) + \mathbf{v}(t)$ [7]. The first equation describes the simplification of the intensities in a set of $n \ll m$ spatial features. The second equation accounts for the evolution of these features over time. $\mathbf{A} \in \mathbb{R}^{n \times n}$ and $\mathbf{C} \in \mathbb{R}^{m \times n}$ are the state-transition and the observation matrices. The two normal random processes $\mathbf{v}(t) \propto \mathcal{N}(0, \mathbf{Q})$, $\mathbf{Q} \in \mathbb{R}^{n \times n}$, and $\mathbf{w}(t) \propto \mathcal{N}(0, \mathbf{R})$, $\mathbf{R} \in \mathbb{R}^{m \times m}$, account for noise in the hidden and in the observed states respectively. Stacking the system for the τ temporal states ($t = 1 \dots \tau$) of \mathcal{S}^p yields the matrix equations $\mathbf{Y}_1^\tau = \mathbf{CZ}_1^\tau + \mathbf{W}$ and $\mathbf{Z}_2^\tau = \mathbf{AZ}_1^{\tau-1} + \mathbf{V}$ where $\mathbf{Z}_p^q = [\mathbf{z}(p), \dots, \mathbf{z}(q)]$ and $\mathbf{Y}_p^q = [\mathbf{y}(p), \dots, \mathbf{y}(q)]$. Neglecting the noise processes, an approximate set of parameters $\mathcal{D}_{\mathbf{x}} = \{\mathbf{A}, \mathbf{C}\}$ is found by constraining \mathbf{C} to be orthogonal [7]. Given the singular value decomposition $\mathbf{Y}_1^\tau = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$, the observation matrix, the hidden variables, and the state-transition matrix read as $\hat{\mathbf{C}} = \mathbf{U}$, $\hat{\mathbf{Z}}_1^\tau = \mathbf{\Sigma}\mathbf{V}^T$, and $\hat{\mathbf{A}} = \hat{\mathbf{Z}}_2^\tau (\hat{\mathbf{Z}}_1^{\tau-1})^\dagger$ respectively. $(.)^\dagger$ denotes the pseudoinverse.

Distance. We calculate at each pixel $\mathbf{x} \in \Omega$ the DT models $\mathcal{D}_{\mathbf{x}}^0$ and $\mathcal{D}_{\mathbf{x}}^p$ for the initial sub-sequence \mathcal{S}^0 and for the current sub-sequence \mathcal{S}^p . In [3], each DT model is represented by the ratio between the norm of the prediction matrix $\|\mathbf{AZ}_1^{\tau-1}\|_2$ and the norm of the residual matrix $\|\mathbf{Z}_2^\tau - \mathbf{AZ}_1^{\tau-1}\|_2$. In this case, the distance between the DT models is defined as the Euclidean distance between the ratios. However, and as illustrated in Fig. 1b, this simplified representation hampers the distinction between the thrombus and collateral vessels, and between the aorta and the nearly static background. We instead maintain the

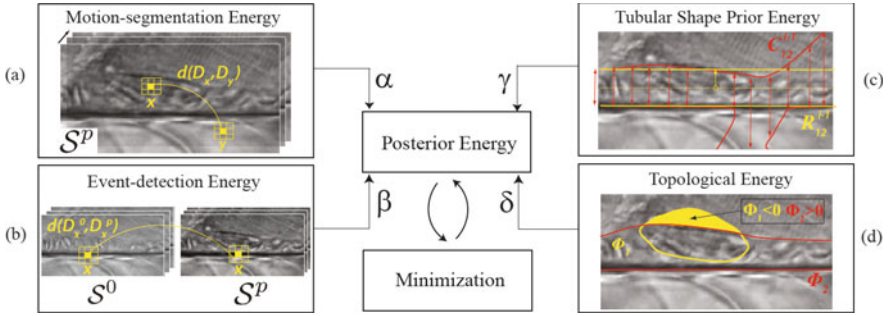


Fig. 2. Proposed framework for joint thrombus-vessel segmentation

complete parametrization of the DT models with their respective matrices \mathbf{A} and \mathbf{C} . The orthogonality constraint on the matrix \mathbf{C} yields the non-linearity of the DT model manifold. A non-linear similarity measure on this manifold thus becomes necessary. Similarly to [8], we choose to employ the Martin distance [10] because of its good precision and of its simple computation.

Motion Segmentation Energy. Analyzing the spatial variations of the DT models at the sub-sequence \mathcal{S}^p is justified by the discrepancy of the motion patterns between the different regions: the thrombus is slowly moving synchronously with the heart beat, the background is nearly static, and the uncoagulated aorta is characterized by the fast and chaotic circulation of blood cells.

Let us assume that the DT models $\{\mathcal{D}_{\mathbf{x}}, \mathbf{x} \in \Omega_k\}$ in each region Ω_k follow a normal distribution $\mathcal{N}(\mu_{\Omega_k}, \sigma_{\Omega_k})$ with respect to the Martin distance d . Extending the probabilistic approach of intensity-based segmentation [5] to dynamic texture segmentation, the *motion segmentation* likelihood energy is defined by

$$E_{ms}(\mathcal{S}^p, \mathcal{S}^0 | \omega) = \sum_{k=1}^3 \sum_{\mathbf{x} \in \Omega_k} \frac{d^2(\mathcal{D}_{\mathbf{x}}, \mu_{\Omega_k})}{2\sigma_{\Omega_k}^2}. \quad (2)$$

Because of the non-Euclidean structure of the DT model space, the mean model μ_{Ω_k} and the standard deviation σ_{Ω_k} cannot be easily computed. The squared Martin distance $d^2(\mathcal{D}_{\mathbf{x}}, \mu_{\Omega_k})$ is thus approximated by its expected value. Given a set of DT model samples $\{\mathcal{D}_{\mathbf{y}}, \mathbf{y} \in \Omega_k^s\}$ computed from a set of randomly selected pixel samples Ω_k^s , the expected value and the standard deviation read as

$$\hat{d}^2(\mathcal{D}_{\mathbf{x}}, \mu_{\Omega_k}) = \frac{1}{|\Omega_k^s|} \sum_{\mathbf{y} \in \Omega_k^s} w_{k,\mathbf{y}} d^2(\mathcal{D}_{\mathbf{x}}, \mathcal{D}_{\mathbf{y}}), \quad \hat{\sigma}_{\Omega_k}^2 = \frac{1}{|\Omega_k^s|} \sum_{\mathbf{x} \in \Omega_k^s} \hat{d}^2(\mathcal{D}_{\mathbf{x}}, \mu_{\Omega_k}). \quad (3)$$

Estimating the standard deviation of the DT model distribution allows us to normalize the likelihood energy. The tractability of Eq. (3) is ensured by choosing $|\Omega_k^s| \ll |\Omega_k|$. To prevent abrupt change of the likelihood energy if a peripheral sample exits the region, the samples in the center of the region are favored by setting $w_{k,\mathbf{y}}$ equal to the Euclidean distance of \mathbf{y} to the boundaries of Ω_k .

Event Detection Energy. In addition to its distinct motion pattern, the thrombus is characterized by its growth in a previously uncoagulated aorta. We model this observation by analyzing the variations of the local motion patterns between the initial sub-sequence \mathcal{S}^0 and the current sub-sequence \mathcal{S}^p .

Let us first consider a pixel \mathbf{x} belonging to the thrombus region in \mathcal{S}^p . This pixel belongs to the uncoagulated aortic region in \mathcal{S}^0 . Its motion pattern is modified. The distance between the DT model $\mathcal{D}_{\mathbf{x}}^0$ observed in \mathcal{S}^0 and the DT model $\mathcal{D}_{\mathbf{x}}^p$ observed in \mathcal{S}^p is thus significant. Please note that the recording is started before the laser injury to ensure that the thrombus has not formed in the initial subsequence \mathcal{S}_0 . Let us then consider a pixel \mathbf{x} belonging to the uncoagulated aorta or to the background region in \mathcal{S}^p . Following the observation that the aorta boundaries remain nearly constant, the pixel \mathbf{x} belongs to the same region in \mathcal{S}^0 . Its motion pattern is unchanged and the distance between $\mathcal{D}_{\mathbf{x}}^0$ and $\mathcal{D}_{\mathbf{x}}^p$ tends to zero. We define the *event-detection* map $\Delta_{\mathbf{x}}^{0,p}$ in each pixel \mathbf{x} as the Martin distance between the DT models $\mathcal{D}_{\mathbf{x}}^0$ and $\mathcal{D}_{\mathbf{x}}^p$. This map is characterized by high values for the thrombus region and by low values for the uncoagulated aorta and the background region.

Assuming a normal distribution $\mathcal{N}(\Delta_{\Omega_k}^{0,p}, \sigma_{\Omega_k}^{0,p})$ of the $\Delta_{\mathbf{x}}^{0,p}$ in the k^{th} region, we encourage the homogeneity of the partition with respect to the temporal changes of the DT models by introducing the *event-detection energy*

$$E_{cd}(\mathcal{S}^p, \mathcal{S}^0 | \omega) = \sum_{k=1}^3 \sum_{\mathbf{x} \in \Omega_k} \frac{(\Delta_{\mathbf{x}}^{0,p} - \Delta_{\Omega_k}^{0,p})^2}{2(\sigma_{\Omega_k}^{0,p})^2}, \text{ with } \Delta_{\mathbf{x}}^{0,p} = d(\mathcal{D}_{\mathbf{x}}^0, \mathcal{D}_{\mathbf{x}}^p). \quad (4)$$

The likelihood energies defined by Eq. (2) and Eq. (4) characterize the spatial and temporal changes of the DT models. However, and as shown in Fig 3c, these energies and the subsequent segmentation results are disrupted by the slow circulation of blood cells in the collateral vessels and by the local non-rigid movements of the background regions. A prior energy thus becomes necessary.

2.2 Tubular Shape Prior Energy

The collateral vessels and the non-rigid movements in the background deform the aorta segmentation towards a non-tubular structure (Fig 3c). We constrain the shape of the aortic region Ω_{12} by introducing a tubular shape prior energy $E_{sp}(\omega)$ into our segmentation framework (Fig 2c). At each iteration i of the energy minimization scheme, we fit a rectangle \mathcal{R}_{12}^{i-1} to the contour (\mathcal{C}_{12}^{i-1}) of the aortic region, taken at the previous iteration. We then define

$$E_{sp}(\omega) = \sum_{k=1}^2 \sum_{x \in \Omega_k} \left[1 - \exp \left(-\frac{d_{R_{12}^{i-1}}^2(\mathbf{x})}{2\sigma^2} \right) \right]. \quad (5)$$

$d_{R_{12}^{i-1}}(\cdot)$ denotes the distance transform of the binary image derived from R_{12}^{i-1} . $\sigma \in \mathbb{R}$ sets the range of feasible deformations. The centerline of R_{12}^{i-1} is approximated by the principal axis of \mathcal{C}_{12}^{i-1} , and its radius by the median distance

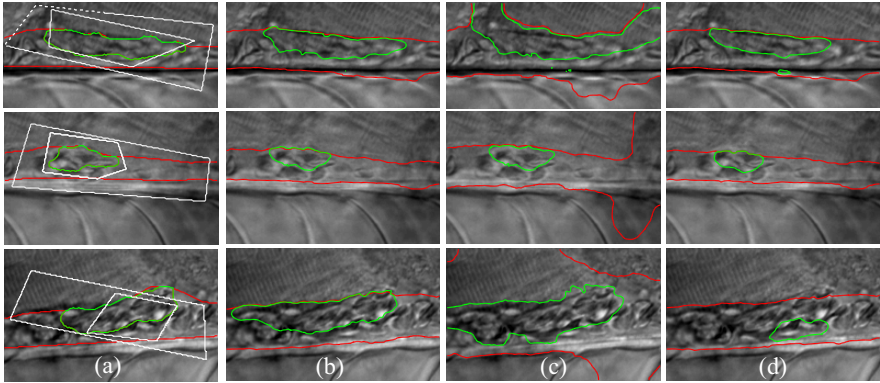


Fig. 3. (a) Ground truth and initialization; (b) Final results with our complete framework; (c)-(d) Results without the shape prior and the event-detection energies. The green contour corresponds to the thrombus, the red to the aorta.

between the boundaries of \mathcal{C}_{12}^{i-1} and the estimated centerline. $E_{sp}(\omega)$ increases with $d_{R_{12}}^2(\mathbf{x})$. It thus prevents the pixels which are far away from the fitted rectangle from belonging to the thrombus (Ω_1) or to the aortic (Ω_2) regions.

2.3 Level-Set Energy Minimization

The three regions $\{\Omega_1, \Omega_2, \Omega_3\}$ are represented by two level-set functions Φ_1 and Φ_2 . We employ the multiphase formulation [14]. We define the thrombus region (Ω_1) by the pixels inside the two contours ($\Phi_1 < 0, \Phi_2 < 0$), the uncoagulated aortic region (Ω_2) by the pixels outside the first contour and inside the second contour ($\Phi_1 > 0, \Phi_2 < 0$), and the background region (Ω_3) by the pixels outside the two contours ($\Phi_1 > 0, \Phi_2 > 0$). The complete aortic region (Ω_{12}) is directly represented by Φ_2 and can thus be easily constrained. As suggested in [1], a topological energy E_{top} is employed to constrain the number of regions in the partition (Fig 21). This energy penalizes the surface area of the remaining region. The weighted sum of the likelihood and prior energies forms the posterior energy

$$E(\omega | \mathcal{S}^p, \mathcal{S}^0) = \alpha E_{ms}(\mathcal{S}^p, \mathcal{S}^0 | \omega) + \beta E_{cd}(\mathcal{S}^p, \mathcal{S}^0 | \omega) + \gamma E_{sp}(\omega) + \delta E_{top}(\omega), \quad (6)$$

which is minimized by alternating the update of the energies with a Laplacian-regularized gradient descent on Φ_1 and Φ_2 [2].

3 Results

Twenty sub-sequences taken from four microscopic video sequences and covering a wide range of biological variations (size of the thrombus, collateral vessels, etc...) were used for testing purpose. The aortic wall is artificially injured by a

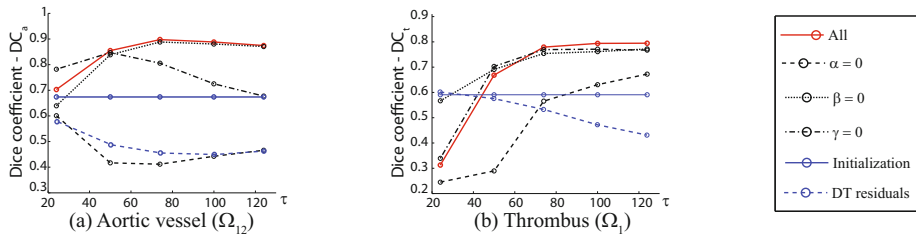


Fig. 4. Mean Dice-coefficient over the 20 sub-sequences, function of τ , α , β , γ

nitrogen-pulse ablating laser. The images are recorded with a Leica DFC420-C camera through a Leica DMRXA Differential Interference Contrast (DIC) - microscope equipped with a $40\times$ water-immersed objective. The ground-truth segmentations are manually delineated by a trained biologist. The same set of empirically chosen parameters ($\alpha = 1$, $\beta = 0.25$, $\gamma = 0.25$, $\delta = 1$) is used over the entire dataset. The dimensions of the DT models, $n = 15$ and $m = 25$, are taken from [4]. Using $|\Omega_k^s| = 10$ samples per region and $\tau = 75$ images of 200×150 pixels per sub-sequences, the computation time on a standard computer (2.00GHz, 8.00GB RAM) is of $15.5 \pm 0.25s$ for DT model learning, of $72.4 \pm 2.05s$ for Martin distance computation, and of $14.5 \pm 5.5s$ for level-set evolution.

The qualitative performance of our method is illustrated in Fig. 3. We display typical segmentation results obtained for $\tau = 75$ if all the introduced energies are employed (b), if the tubular shape constraint is dismissed (c), and finally if the event-detection likelihood is discarded (d).

The accuracy of our algorithm is quantitatively evaluated for increasing values of τ with an overlap measure, the Dice-coefficient, for both the aortic (DC_a) and the thrombus (DC_t) regions (cf. Fig. 4). While a minimal window size is necessary to adequately capture the motion patterns, choosing the smallest possible τ prevents the temporal smoothing of the segmentation results. To this regards, $\tau = 75$ is a good compromise value. Our approach yields mean DC values of $DC_a \geq 0.90$ and $DC_t \geq 0.78$. These respectively fall down to $DC_a = 0.47$ (-48%) and $DC_t = 0.57$ (-27%) if the DT residual energy (Fig. 1) is employed instead of the proposed energies. This shows the better performance of our method in comparison to our benchmark approach [3]. We quantitatively illustrate the importance of incorporating the event-detection and the shape prior energies in addition to the motion-segmentation energy. Discarding the event-detection energy ($\beta = 0$) decreases the mean DC_t to 0.75 (-4%), and canceling the shape prior probability ($\gamma = 0$) yields a mean DC_a of 0.80 (-10%). We moreover observe in Fig. 4a that the longer the subsequence is, the more likely the perturbations in the background become. This makes the shape prior energy increasingly necessary for $\tau \geq 75$. The motion segmentation energy remains essential to obtain accurate segmentation results. If not accounted ($\alpha = 0$), the mean DC_a decreases to 0.41 (-63%) and the mean DC_t to 0.57 (-27%).

4 Conclusion

In this work, we propose an innovative solution to the complex problem of joint thrombus and aortic vessel segmentation in low contrast, cluttered, and highly dynamic microscopic video sequences. By characterizing the spatial and temporal discrepancy of the motion patterns with two dynamic texture-based likelihood energies, and by constraining the aortic region with a shape constraint, we model the major facets of our biomedical imaging setting. Though applied to microscopy, we believe that the capacity of the proposed algorithm to analyze complex dynamic scenes could be beneficial to other medical imaging modalities.

References

1. Baust, M., Navab, N.: A spherical harmonics shape model for level set segmentation. In: Daniilidis, K. (ed.) ECCV 2010, Part III. LNCS, vol. 6313, pp. 580–593. Springer, Heidelberg (2010)
2. Baust, M., Zikic, D., Navab, N.: Diffusion-based regularisation strategies for variational level set segmentation. In: BMVC, pp. 19.1–19.11. BMVA Press (2010)
3. Brieu, N., Serbanovic-Canic, J., Cvejic, A., Stemple, D., Ouwehand, W., Navab, N., Groher, M.: Thrombus segmentation by texture dynamics from microscopic image sequences. In: SPIE Medical Imaging (2010)
4. Chan, A., Vasconcelos, N.: Modeling, clustering, and segmenting video with mixtures of dynamic textures. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 30, 909–926 (2008)
5. Cremers, D., Rousson, M., Deriche, R.: A review of statistical approaches to level set segmentation: Integrating color, texture, motion and shape. *Intl. J. Computer Vision* 72, 195–215 (2007)
6. Dong, G., Ray, N., Acton, S.: Intravital leukocyte detection using the gradient inverse coefficient of variation. *IEEE Trans. Med. Imag.* 24(7), 910–924 (2005)
7. Doretto, G., Chiuso, A., Wu, Y., Soatto, S.: Dynamic textures. *Intl. J. Computer Vision* 51, 91–109 (2003)
8. Doretto, G., Cremers, D., Favaro, P., Soatto, S.: Dynamic texture segmentation. In: ICCV 2003, vol. 2, pp. 1236–1242 (2003)
9. Li, F., Zhou, X., Zhao, H., Wong, S.: Cell segmentation using front vector flow guided active contours. In: Yang, G., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5762, pp. 609–616. Springer, Heidelberg (2009)
10. Martin, R.: A metric for ARMA processes. *Trans. on Signal Processing* 48(4) (2002)
11. O'Connor, M., Salles, I., Cvejic, A., Watkins, N., Ouwehand, W.: Functional genomics in zebrafish permits rapid characterization of novel platelet membrane proteins. *Blood* 113(19), 4754–4762 (2009)
12. Palaniappan, K., Ersoy, I., Nath, S.: Moving object segmentation using the flux tensor for biological video microscopy. In: Ip, H., Au, O., Leung, H., Sun, M., Ma, W., Hu, S. (eds.) PCM 2007. LNCS, vol. 4810, pp. 483–493. Springer, Heidelberg (2007)
13. Schmutge, S., Keller, S., Nguyen, N., Souvenir, R., Huynh, T., Clemens, M., Shin, M.: Segmentation of vessels cluttered with cells using a physics based model. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 127–134. Springer, Heidelberg (2008)
14. Vese, L., Chan, T.: A multiphase level set framework for image segmentation using the Mumford and Shah model. *Intl. J. on Computer Vision* 50(3), 271–293 (2002)

3D Kidney Segmentation from CT Images Using a Level Set Approach Guided by a Novel Stochastic Speed Function

Fahmi Khalifa¹, Ahmed Elnakib¹, Garth M. Beache², Georgy Gimel'farb³,
Mohamed Abo El-Ghar⁴, Rosemary Ouseph⁵, Guela Sokhadze¹,
Samantha Manning¹, Patrick McClure¹, and Ayman El-Baz^{1,*}

¹ BioImaging Laboratory, Bioengineering Department, University of Louisville,
Louisville, KY, USA

² Diagnostic Radiology Department, University of Louisville, Louisville, KY, USA

³ Department of Computer Science, University of Auckland, Auckland, New Zealand

⁴ Urology and Nephrology Department, University of Mansoura, Mansoura, Egypt

⁵ Department of Medicine, University of Louisville, Louisville, KY, USA

Abstract. Kidney segmentation is a key step in developing any non-invasive computer-aided diagnosis (CAD) system for early detection of acute renal rejection. This paper describes a new 3-D segmentation approach for the kidney from computed tomography (CT) images. The kidney borders are segmented from the surrounding abdominal tissues with a geometric deformable model guided by a special stochastic speed relationship. The latter accounts for a shape prior and appearance features in terms of voxel-wise image intensities and their pair-wise spatial interactions integrated into a two-level joint Markov-Gibbs random field (MGRF) model of the kidney and its background. The segmentation approach was evaluated on 21 CT data sets with available manual expert segmentation. The performance evaluation based on the receiver operating characteristic (ROC) and Dice similarity coefficient (DSC) between manually drawn and automatically segmented contours confirm the robustness and accuracy of the proposed segmentation approach.

1 Introduction

Kidney segmentation from abdominal CT images is an essential step for many high-level processing tasks, such as localization of pathology, radiotherapy planning, and computer-integrated surgery. However, due to image noise, acquisition artifacts, gray level inhomogeneities, and similar visual appearances of adjacent structures, accurate segmentation of the kidney still remains a challenge [1]. In recent years, many automated and semi-automated approaches have been developed to address these challenges. In particular, Pohle and Toennies [2] developed an automatic region-growing algorithm for segmenting anatomical structures. Their approach estimated the homogeneity criterion from the characteristics of the images to be segmented. However, due to gray level similarities

* Corresponding author:- Tel.: (502)-852-5092, E-mail: aselba01@louisville.edu

between the kidney and the abdominal tissues, as well as sensitivity to initial seed point locations, region growing-based approaches are not effective for CT kidney segmentation.

Wang et al. [3] proposed a constrained optimization deformable contour method in which the degree of contour interior homogeneity is computed as an extra constraint within the level set energy minimization framework. However, deformable models fail in the case of noise, spurious edges, poor image resolution, diffused boundaries, or occluded shapes if they do not include a *priori* knowledge to constrain the contour deformation within an admissible range. Tsagaan et al. [4] proposed an automatic deformable model-based approach to segment the kidney from CT images. Their method is based on a non-uniform rational B-spline surface representation and statistical shape information of the kidney, which is then incorporated into the objective function as an additional energy term. Their results, evaluated on 33 CT images, seem reasonable in some cases, but differ markedly from the manually segmented results in others.

Recently, Huang et al. [5] proposed a multiphase level set approach with multi-dynamic shape models to segment the kidneys on abdominal CT images. Spiegel et al. [6] proposed a kidney segmentation framework based on the active shape model (ASM) that was combined with a curvature-based non-rigid registration approach to solve the point correspondence problem of the training data. In general, knowledge-based approaches are computationally intensive, and their accuracy depends on the size of the training data. Freiman et al. [7] proposed a model-based kidney segmentation approach from CT images based on maximum a posteriori-Markov random field (MAP-MRF) estimation of the current image. The MAP-MRF estimation is obtained by using the graph min-cut technique. Campadelli et al. [8] proposed an automatic, gray-level based segmentation framework based on a multiplanar fast marching method. Their segmentation performance was evaluated based only on visual assessment.

To overcome the aforementioned limitations, we introduce a 3-D extension of our previous 2-D stochastic guiding force presented in [9] to guide the evolution of a 3-D geometric deformable model to extract the kidney region from CT images. The proposed 3-D stochastic guiding force accounts for a 3-D shape prior, 1st-order intensity model, and a 3-D 2nd-order spatial interaction model between the kidney voxels and its background.

2 The Proposed Level Set-Based Segmentation Approach

In recent years, level set-based deformable models have been applied to medical image segmentation with considerable success because of the flexibility of the evolving boundary and the lack of need for parameterizations. The object-background boundary at each moment t is represented by a zero level $\phi_t(x, y, z) = 0$ of an implicit level set function, namely a distance map $\phi_t(x, y, z)$ of the signed minimum Euclidean distances from every voxel to the boundary (negative for interior and positive for exterior voxels). The distance map is evolved iteratively with the evolution being guided by a speed function $V_n(x, y, z)$ [10]:

$$\phi_{n+1}(x, y, z) = \phi_n(x, y, z) - \tau V_n(x, y, z) |\nabla \phi_n(x, y, z)| \quad (1)$$

where n indicates the time instant $t = n\tau$ (taken with a step $\tau > 0$), and $\nabla \phi_n = [\frac{\partial \phi_n}{\partial x}, \frac{\partial \phi_n}{\partial y}, \frac{\partial \phi_n}{\partial z}]$ is the gradient of $\phi_n(x, y, z)$. Conventional speed functions accounting for image intensities, object edges, gradient vector flow, etc., are unsuccessful on very noisy images with low object-background intensity gradients. The results are improved by involving shape priors (e.g., [5,6,11]). To obtain more accurate results, our stochastic speed function accounts for both the shape prior and appearance features associated with image intensities and their spatial interactions integrated into a 3-D two-level joint MGRF model.

Shape-Appearance Guided Evolution: Let $\mathbf{Q} = \{0, 1, \dots, Q-1\}$ and $\mathbf{L} = \{0, 1\}$ denote a finite set of integer gray values and a binary set of object (“1”) and background (“0”) labels, respectively. Let \mathbf{R} denote a 3-D arithmetic (x, y, z) -lattice that supports a given grayscale CT data $\mathbf{g} : \mathbf{R} \rightarrow \mathbf{Q}$ to be segmented and its goal binary “object - background” region map $\mathbf{m} : \mathbf{R} \rightarrow \mathbf{L}$. The image \mathbf{g} , being co-aligned to a shape prior, and its map \mathbf{m} are described with a joint probability model $P(\mathbf{g}, \mathbf{m}) = P(\mathbf{g}|\mathbf{m})P(\mathbf{m})$ combining a 2^{nd} -order MGRF $P(\mathbf{m})$ of region labels with the shape prior and a conditionally independent random field $P(\mathbf{g}|\mathbf{m})$ of image intensities given the map. The map model $P(\mathbf{m}) = P_s(\mathbf{m})P_h(\mathbf{m})$ has two parts: (i) a shape prior probability $P_s(\mathbf{m})$ and (ii) a 2^{nd} order MGRF model $P_h(\mathbf{m})$ of a spatially homogeneous region map \mathbf{m} for the image \mathbf{g} .

Conditional Intensity Model: To build an initial map \mathbf{m} , we need to estimate the marginal intensities distribution for the object (i.e., kidney) and background. The empirical gray level distribution, $P(\mathbf{g}) = \prod_{(x,y,z) \in \mathbf{R}} P_{\text{MIX}}(g_{x,y,z})$, is separated into object and background components, $(p(q|\lambda) : q \in \mathbf{Q}; \lambda \in \mathbf{L})$, by close approximation with a linear combination of discrete Gaussians (LCDG) [1], a modified version of our previous linear combination of continuous Gaussians probabilistic model [12]. This approximation adapts the segmentation to changing appearance, such as non-linear intensity variations caused by patient weight and data acquisition system (scanner type and scanning parameters). The LCDG separates each factor of the empirical gray level distribution more accurately than conventional mixtures of only positive Gaussians, thus yielding a better initial region map formed by voxel-wise classification of the CT images gray values.

Spatial Voxel Interaction Model: To smooth the evolution of the level set and get more accurate segmentation, spatially homogeneous interactions between the region labels are modeled with a generic MGRF of a region map that accounts only for 3-D voxelwise interactions between each region label and its characteristic neighborhood. For simplicity, we restrict the interaction structure to the nearest voxel 26-neighbors (Fig. 1). By symmetry considerations, we assume that the potentials are independent of the relative orientation of each voxel pair and depend only on whether the labels are equal or not. Under these

¹ For complete details, please see: <https://louisville.edu/speed/bioengineering/faculty/bioengineering-full/dr-ayman-el-baz/supplemental-materials>

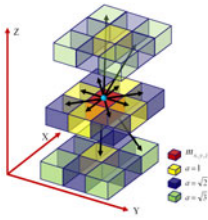


Fig. 1. Pairwise voxel interaction for the 26 neighbors in a 3-D MGRF image model

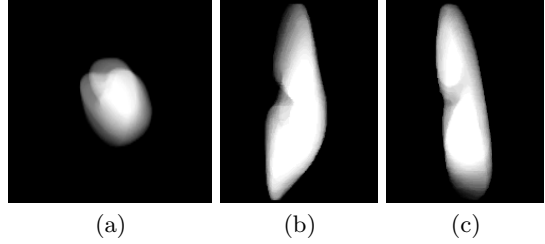


Fig. 2. 3D kidney shape model projected onto 2D axial (a), coronal (b), and sagittal (c) planes for visualization

restrictions, it is a 3-D extension of the conventional auto-binomial, or Potts model, differing only in that the potentials are estimated analytically.

The 26-neighborhood has three types of symmetric pairwise interactions specified by the absolute distance a between two voxels in the same and adjacent CT slices ($a = 1, \sqrt{2}$, and $\sqrt{3}$, respectively): (i) the closest pairs with the inter-voxel coordinate offsets $\mathbf{N}_1 = \{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$; (ii) the farther diagonal pairs with the offsets $\mathbf{N}_{\sqrt{2}} = \{(0, 1, \pm 1), (1, 0, \pm 1), (1, \pm 1, 0)\}$; and (iii) the farthest diagonal pairs with the offsets $\mathbf{N}_{\sqrt{3}} = \{(1, \pm 1, \pm 1)\}$. The potentials of each type are bi-valued because only the coincidence of the labels is taken into account: $\mathbf{V}_a = \{v_{a,\text{eq}}; v_{a,\text{ne}}\}$ where $v_{a,\text{eq}} = \gamma$ if $l = l'$ and $v_{a,\text{ne}} = -\gamma$ if $l \neq l'$; $a \in \mathbf{A} = \{1, \sqrt{2}, \sqrt{3}\}$. Let $\mathbf{N} = \{\mathbf{N}_a, a \in \mathbf{A}\}$ and $f_{a,\text{eq}}(\mathbf{m})$ denote the relative frequency of the equal label pairs in the equivalent voxel pairs $\{((x, y, z), (x+\xi, y+\eta, z+\zeta)): (x, y, z) \in \mathbf{R}; (x+\xi, y+\eta, z+\zeta) \in \mathbf{R}; (\xi, \eta, \zeta) \in \mathbf{N}\}$, then the MGRF model is:

$$P_h(\mathbf{m}) \propto \exp \sum_{(x,y,z) \in \mathbf{R}} \sum_{(\xi,\eta,\zeta) \in \mathbf{N}} \mathbf{V}_a(m_{x,y,z}, m_{x+\xi,y+\eta,z+\zeta}) \quad (2)$$

The initial region map results in the approximate analytical maximum likelihood estimates of the potentials²: $v_{a,\text{eq}} = -v_{a,\text{ne}} \approx 2f_{\text{eq}}(\mathbf{m}) - 1$; that allow for computing the voxel-wise probabilities $p_{h:x,y,z}(m_{x,y,z} = \lambda)$ of the labels; $\lambda \in \mathbf{L}$, at each step of the boundary evolution.

Probabilistic Shape Prior: To enhance the segmentation accuracy, the expected shape of the goal object is constrained with a probabilistic shape prior (Fig. 2). A training database collected from different subjects are co-aligned by rigid, namely affine, 3-D transformations maximizing their mutual information (MI) [13]. The shape prior is a spatially-variant independent random field of region labels $P_s(\mathbf{m}) = \prod_{(x,y,z) \in \mathbf{R}} p_{s:x,y,z}(m_{x,y,z})$ for the co-aligned, manually segmented training CT images. The factors are the empirical voxel-wise object

² For complete proof, please see: <https://louisville.edu/speed/bioengineering/faculty/bioengineering-full/dr-ayman-el-baz/supplemental-materials>

$p_{s:x,y,z}(1)$, and background $p_{s:x,y,z}(0) = 1 - p_{s:x,y,z}(1)$ probabilities. Each CT data to be segmented is initially co-aligned to the training database.

Stochastic Speed Function: All of the above models contribute to the voxel-wise guidance of the evolving level set. Let κ be the mean contour curvature. Let $\vartheta(x, y, z)$ specify the evolution magnitude and direction:

$$\vartheta(x, y, z) = \begin{cases} -P_{1:x,y,z} & \text{if } P_{1:x,y,z} > P_{0:x,y,z} \\ P_{0:x,y,z} & \text{otherwise} \end{cases} \quad (3)$$

where $P_{1:x,y,z} = \frac{\Omega_{1:x,y,z}}{\Omega_{1:x,y,z} + \Omega_{0:x,y,z}}$ and $P_{0:x,y,z} = 1 - P_{1:x,y,z}$. Here, $\Omega_{1:x,y,z} = p(q|1)p_{h:x,y,z}(1)p_{s:x,y,z}(1)$; $\Omega_{0:x,y,z} = p(q|0)(1 - p_{h:x,y,z}(1))(1 - p_{s:x,y,z}(1))$, and $p_{h:x,y,z}(1)$ is the probability of the object label in the Potts model $P_h(\mathbf{m})$. The stochastic speed function in Eq. (3) is defined as in [9]: $V(x, y, z) = \kappa\vartheta(x, y, z)$. The steps of our segmentation approach are summarized in Algorithm 1.

Algorithm 1. Key Steps for Level Set Segmentation

1. Construct the probabilistic shape prior from the training data sets.
 2. Approximate the empirical gray level distribution by using the LCDG with two dominant Gaussian modes.
 3. Form an initial region map \mathbf{m} using the estimated LCDG models.
 4. Find the Gibbs potentials for the MGRF model from the initial map.
 5. Find the stochastic speed function defined in Eq. (3).
 6. Evolve the level set using the determined speed function.
-

3 Experimental Results

The proposed approach has been tested on *in-vivo* 3-D CT kidney data sets (7 for training and 14 for testing and performance evaluation). The images were acquired by a GE light speed plus scanner (General Electric, Milwaukee, USA) using the following parameters: 120 KV, 250 mA, slice thickness of 0.9 mm, and FOV of 360 mm. To minimize the effect of inter-observer variations, each CT data is segmented by three independent experts and the “ground truth” is considered as the common segmented part of their segmentations. The CT images have marginal intensity distributions that are mixtures of two dominant modes: one mode for the kidney object and another mode for the background. Basic density estimation steps using the LCDG models are illustrated in Fig. 3 and Fig. 5 demonstrates the results of kidney segmentation.

The performance of our segmentation approach is evaluated based on the voxel-based overlap measured by using the Dice similarity coefficient (DSC) [14] and the average perpendicular distance (APD) between the automatic segmentation (\mathbf{C}) and the ground truth (\mathbf{G}). To measure the distances, one has to accurately co-locate the point-to-point correspondences between the borders. In

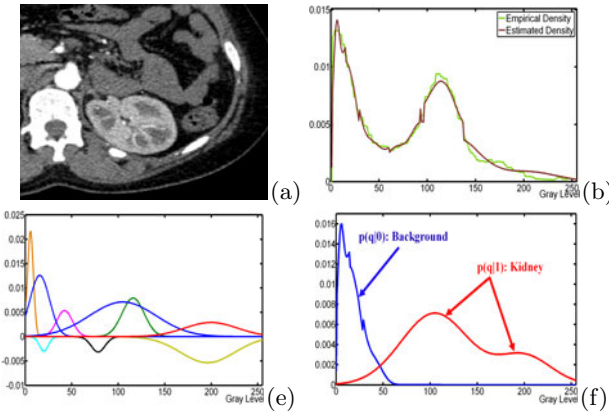


Fig. 3. LCDG-based modeling of the marginal probability distribution: (a) typical CT image, (b) its gray level density estimation (the final estimated density (brown) for the empirical density (green)), (c) sign-alternate LCDG components, and (d) final estimated LCDG for each class

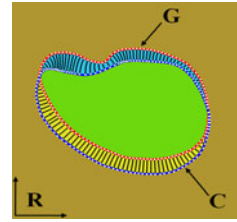


Fig. 4. 2-D illustration of the correspondences found by the solution of the Laplace equation between the ground truth \mathbf{G} (manual expert segmentation) and automatic segmentation \mathbf{C} obtained by the proposed segmentation approach

this paper, the correspondences, or matches between the borders' points (see e.g., Fig. 4), are found by solving the Laplace equation:

$$\nabla^2\psi = \frac{\partial^2\psi}{\partial^2x^2} + \frac{\partial^2\psi}{\partial^2y^2} + \frac{\partial^2\psi}{\partial^2z^2} = 0 \tag{4}$$

for a scalar potential field ψ . Table 1 represents the DSC and APD statistics obtained for all test data sets.

To highlight the advantages of our approach, we compared it to the shape-based approach proposed in [11] based on the average volumetric error $E_r = \frac{1}{\mathbf{G}} (FP + FN)$, where FP and FN are the false positive and false negative segmentation errors, respectively. Table 2 compares the segmentation results over all the test data sets for our approach and the shape-based approach proposed in [11] with respect to the radiologist's segmentation. Differences between the average volumetric errors for our approach and the shape-based approach [11] are statistically significant by the unpaired t -test (the two-tailed P -value is $\leq 10^{-4}$). The final 3-D kidney segmentations for two of the test data sets and their associated FP and FN errors are shown in Fig. 6.

Another major metric to test the performance of our segmentation approach is to compute the receiver operating characteristic (ROC). Each point on the graph is generated by using a different cut point (i.e., classification threshold). Figure 7 shows the ROC curves of our three feature approach, intensity-based only (I), the intensity and spatial interactions-based ($I + S$), and the intensity and shape-based ($I + P$) level set segmentation. The figure clearly demonstrates that the area under the ROC curve is greatest for our approach ($A_z = 0.9423$).

Table 1. Segmentation accuracies over all test data sets. Note that APD values are in mm.

	DSC	APD
Min.	0.950	0.00
Max.	0.993	3.12
Mean	0.970	1.25
Std.	0.019	0.68

Table 2. Comparative segmentation accuracy over all test data sets.

	Algorithm	
$E_r, \%$	Our	[11]
Min.	1.26	6.76
Max.	7.72	20.08
Mean	3.71	13.95
Std.	2.18	5.96

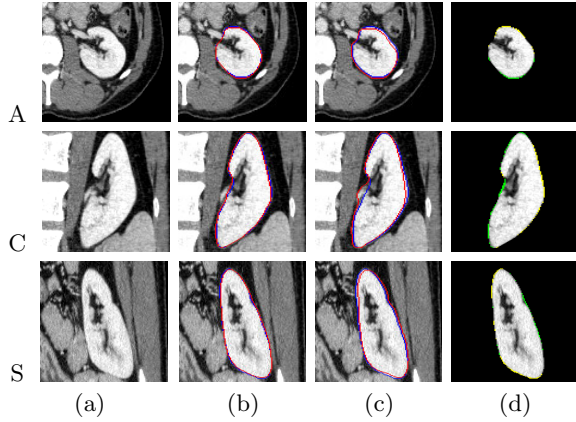


Fig. 5. 3-D kidney segmentation projected onto 2-D axial (A), coronal (C), and sagittal (S) planes for visualization: (a) 2-D CT images, (b) our segmentation (red) compared with the ground truth (blue), (c) the segmentation with the algorithm in [11], and (d) the associated FP (green) and FN (yellow) errors of our segmentation w.r.t. the ground truth

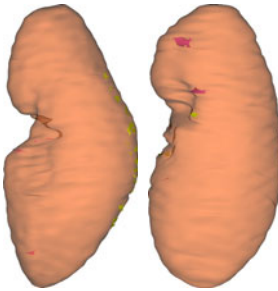


Fig. 6. 3-D visualization for the segmented kidneys using the proposed segmentation approach for two of the test data sets and their associated FP (pink) and FN (yellow) errors

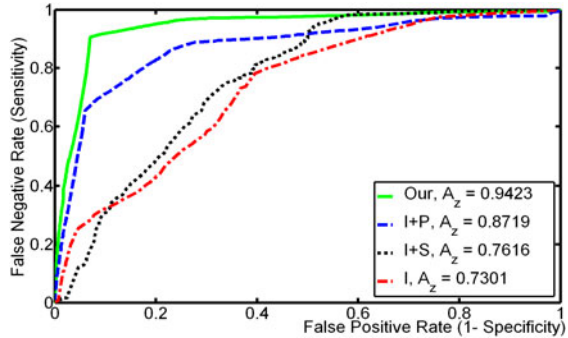


Fig. 7. The ROC curves for different level set-based segmentation guided by: the intensity information only (I ; red), the combined intensity and spatial interactions features ($I + S$; black), the combined intensity and shape features ($I + P$; blue), and the integrated three features (our; green). Note that A_z stands for the area under the curve.

4 Conclusions

This paper has presented a novel and automated 3-D approach for the segmentation of the kidney from abdominal CT images. Incorporation of the CT images'

features, namely the 3-D probabilistic shape, the 1st-order intensity, and the 2nd-order spatial interaction into the speed function notably improves the level set evolution and increases the segmentation accuracy and robustness based on both ROC curves and Dice similarity coefficient metrics. Experimental results showed that the proposed segmentation approach outperformed other methods for kidney segmentation. We plan to ultimately include this segmentation method in a kidney-dedicated CAD system designed for the early detection of acute renal rejection and treatment planning.

References

1. Pham, D.L., Xu, C., Prince, J.L.: Current Methods in Medical Image Segmentation. *Annu. Rev. Biomed. Eng.* 2, 315–338 (2000)
2. Pohle, R., Toennies, K.D.: A New Approach for Model-based Adaptive Region Growing in Medical Image Analysis. In: Skarbek, W. (ed.) CAIP 2001. LNCS, vol. 2124, pp. 238–246. Springer, Heidelberg (2001)
3. Wang, X., He, L., Wee, W.: Deformable Contour Method: A Constrained Optimization Approach. *Int. J. Comput. Vision* 59(1), 87–108 (2004)
4. Tsagaan, B., Shimizu, A., Kobatake, H., Miyakawa, K.: An Automated Segmentation Method of Kidney Using Statistical Information. In: Dohi, T., Kikinis, R. (eds.) MICCAI 2002. LNCS, vol. 2488, p. 556. Springer, Heidelberg (2002)
5. Huang Y.-P., Chung, P.-C., Huang, C.-L., Huang, C.-R.: Multiphase Level Set with Multi Dynamic Shape Models on Kidney Segmentation of CT Image. In: IEEE Biomedical Circuits and Systems Conference (BioCas 2009), pp. 141–144 (2009)
6. Spiegel, M., Hahn, D.A., Daumb, V., Waszab, J., Hornegger, J.: Segmentation of Kidneys Using a New Active Shape Model Generation Technique based on Non-rigid Image Registration. *Comput. Med. Imag. Graphics* 33, 19–39 (2009)
7. Freiman, M., Kronman, A., Esses, S., Joskowitz, L., Sosna, J.: Non-parametric Iterative Model Constraint Graph min-cut for Automatic Kidney Segmentation. In: Jiang, T., Navab, N., Pluim, J., Viergever, M. (eds.) MICCAI 2010. LNCS, vol. 6363, pp. 73–80. Springer, Heidelberg (2010)
8. Campadelli, P., Casiraghi, E., Pratisoli, S.: A Segmentation Framework for Abdominal Organs from CT Scans. *Artif. Intell. Med.* 50, 3–11 (2010)
9. Khalifa, F., El-Baz, A., Gimel'farb, G., Ousephand, R., Abu El-Ghar, M.: Shape-Appearance Guided Level Set Deformable Model for Image Segmentation. In: International Conference on Pattern Recognition (ICPR 2010), pp. 4581–4584 (2010)
10. Osher, S., Fedkiw, R.: *Level Set Methods and Dynamic Implicit Surfaces*. Springer, New York (2006)
11. Tsai, A., Yezzi Jr., A., Wells, W., Tempany, C., Tucker, D., Fan, A., Grimson, W.E., Willsky, A.: A Shape-based Approach to Curve Evolution for Segmentation of Medical Imagery. *IEEE Trans. Med. Imag.* 22(2), 137–154 (2003)
12. Farag, A., El-Baz, A., Gimel'farb, G.: Precise Segmentation of Multimodal Images. *IEEE Trans. Image Process.* 15(4), 952–968 (2006)
13. Viola, P., Wells III, W.M.: Alignment by Maximization of Mutual Information. *Int. J. Comput. Vision* 24(2), 137–154 (1997)
14. Dice, L.R.: Measures of the Amount of Ecologic Association Between Species. *Ecological Society of America* 26(3), 297–302 (1945)

Mixture of Segmenters with Discriminative Spatial Regularization and Sparse Weight Selection*

Ting Chen¹, Baba C. Vemuri^{1,**},
Anand Rangarajan¹, and Stephan J. Eisenschenk²

¹ Department of CISE, University of Florida, Gainesville, FL, USA

{tichen,vemuri,anand}@cise.ufl.edu

² Department of Neurology, University of Florida, Gainesville, FL, USA

stephan.eisenschenk@neurology.ufl.edu

Abstract. This paper presents a novel segmentation algorithm which automatically learns the combination of weak segmenters and builds a strong one based on the assumption that the locally weighted combination varies w.r.t. *both* the weak segmenters and the training images. We learn the weighted combination during the training stage using a discriminative spatial regularization which depends on training set labels. A closed form solution to the cost function is derived for this approach. In the testing stage, a sparse regularization scheme is imposed to avoid overfitting. To the best of our knowledge, such a segmentation technique has never been reported in literature and we empirically show that it significantly improves on the performances of the weak segmenters. After showcasing the performance of the algorithm in the context of atlas-based segmentation, we present comparisons to the existing weak segmenter combination strategies on a hippocampal data set.

1 Introduction

Brain MR image analysis and its associated application in the diagnosis and treatment of brain-based diseases has attracted immense attention in the past two decades. The segmentation of brain neuroanatomy is one of the key steps in medical image analysis. For example, researchers are interested in the study of hippocampal structures due to the critical role they play in many neuro-disorders including dementia, epilepsy and schizophrenia. In order to avoid the tedium involved in manual segmentation, a technique that is able to automatically segment the hippocampi from 3D brain MR scans is of great clinical interest.

Several techniques have been proposed in the literature to segment the hippocampus. One direct approach in [1] seeks to build an atlas from the training images with manual labels and deform it to the test image using a deformable

* This research was in part supported by funding from NIH NS046812 to BCV and the Wayne Densch Epilepsy Research Fund to SJE.

** Corresponding author.

registration. However, it has limitations in accuracy for segmenting small structures like the hippocampus with low contrast intensity boundaries. Alternatively, there exists a class of methods that achieve more robustness and full automation by extracting features from the image at each voxel and resorting to machine learning techniques to label the voxel. In [2], Golland *et al.* use the support vector machine (SVM) to classify the features chosen by PCA from a large feature set. In [3], Morra *et al.* adopt Adaboost to select the features and resort to the SVM for classification. Both frameworks obtain relatively good approximation to the boundary of the hippocampus and the feature selection method simplifies the experts' effort in choosing informative features from a large feature set. However, one still needs to develop a very large feature pool containing potentially useful features, a non-trivial task in general.

Recently, it was shown that combining multiple atlas based segmentations improves the segmentation accuracy [4,5]. As one of the most popular combination strategies, majority voting was shown to improve the accuracy and robustness of weak hypotheses. In [4], Artaechevarria *et al.* propose an image segmentation algorithm that combines multiple atlas-based segmenters based on weighted voting with the weights estimated from the local similarity between each atlas and the test image. Due to this specific design, the weak segmenters for this algorithm are limited to atlas-based segmentation. Besides, in order for the voting based technique to work well, a sufficiently large number of atlases are needed and a robust registration algorithm is required so as to have a relatively accurate segmentation for each atlas. A different combination strategy called SuperDyn was proposed by Khan *et al.* [5], wherein supervised learning was used for computing the weighted combination. Subsequently, dynamic information based on registration accuracy between the atlas and the test image was employed for the weight selection. SuperDyn *independently* estimates the weights for each weak segmenter at each voxel. However, this is clearly inadequate since it has been known for long that strong spatial dependencies exist in most real images [6]. Moreover, the aforementioned dynamic selection also restricts this technique to atlas-based weak segmenters.

In this paper, we propose a novel segmentation algorithm dubbed SegMix which is different from *all* of the aforementioned frameworks. Spatial smoothness and boundary discontinuities in the anatomical structures are explicitly incorporated into a discriminative regularizer in the training stage, resulting in a general technique capable of utilizing a vast variety of weak segmenters. SegMix assumes that the combination weights depend not only on the weak learners but also on the training data. *This is analogous to a medical consultation carried out by a group of doctors on a number of patients. It is well-justified to assume that each patient's personal condition has a different effect on the experts' final decision.* We carefully treat the problem of overfitting (which can occur from having too many weights) by utilizing the previously mentioned spatial weight regularization and via a novel non-parametric testing sieve. This makes SegMix substantially and thematically different from the other combination strategies, e.g. SuperDyn and Voting, which basically assume that the weights *only vary*

w.r.t. the weak segmenters. Due to this novel aspect of our framework, combining a very small number of weak segmenters can lead to dramatically significant improvements which is validated in our experiments. Furthermore, a novel scheme is imposed in the testing phase, wherein a strong segmenter is constructed by approximating the test image with a *sparse* combination of the training data and only choosing the learned weights corresponding to those sparsely selected training images. Intuitively speaking, in the training stage, we use a *cooperation* mechanism on the weak segmenters so as to achieve the best segmentation for each training image, while in the testing phase, we use a *competition* mechanism to select only the relevant images from the training set for a particular test image. The result is an algorithm driven by “*co-opetition*” which uses the previously learned cooperation skills of the competitively selected training data to let the weak segmenters collaborate and obtain a strong segmentation. *Note that as more expert driven manual delineations become available, they can be used as weak segmenters in our framework.*

2 Methodology

2.1 The Segmentation Mixture Setup

In this section, we present the methodology of our algorithm. We begin with illustrating the basic framework of the technique in Fig. 1. In the training stage, the optimal locally weighted combination of the weak segmentations are estimated to best approximate the ground truth label for each training image. As shown in the figure, the weight matrix W_{nt} is associated with the n^{th} training image and t^{th} weak segmenter. Since we compute local weights, the spatial interactions among voxels are modeled closely following the Discriminative Random Field

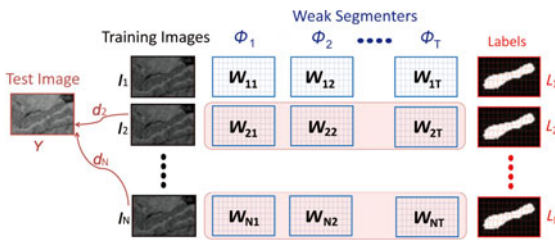


Fig. 1. Framework of the proposed algorithm

(DRF) [6]. In the testing stage, we compute the sparse combination of the training images to approximate the test image and only the learned weights associated with those selected training images are used to construct the strong segmenter for that particular test image. The figure shows an example of picking I_2 and I_N as the sparse representation of the test image and only the weights associated with them, W_{21}, \dots, W_{2T} and W_{N1}, \dots, W_{NT} , are used to construct the final strong segmenter for test image Y .

Before getting into the details of the algorithm, we first introduce two key terms that will be used in the rest of this paper.

Signed Distance Functions (SDF) are used to represent the shape of the structure in the segmentation. Each weak segmentation output is represented by a SDF and the locally weighted combination of them corresponds to a strong segmenter. Several techniques exist in literature for linear combination of SDFs. In [7], Pohl *et al.* embed SDFs into the linear space of LogOdds, where addition and scalar multiplication are in closed form, while in [8] an SDF is mapped to the square-root density space via a Schrödinger wave function, where a variety of Riemannian operations are easily computed. However, we will resort to a simpler idea pointed out by Leventon *et al.* in [9]. They claim that using the signed distance transform shape representation is tolerant to slight misalignment and hence a rough alignment of the data during pre-processing will avoid solving for the general correspondence problem in SDF combination.

A **Neighborhood Graph \mathbf{G}** is computed for each training data to store the spatial interactions and dependencies between voxels. Given a training sample, let $L(x)$ be the signed distance transform of the label image and M be the number of voxels. The $M \times M$ -dimensional neighborhood graph matrix G is computed from the following DRF formulation: $G(i, j) = \exp(-\|(L(x_i) - L(x_j))\|_2^2)$, where $j \in \mathcal{N}(i)$. In 3D, the neighborhood region of the i^{th} voxel $\mathcal{N}(i)$ is computed through 6, 18 or 26-connectivity. This neighborhood graph will be used to gate the distance between the weights corresponding to the i^{th} and j^{th} voxels.

2.2 Training Stage: Discriminative Spatial Weight Regularization

Our segmentation algorithm takes as input a set of weak segmentation results and combines them via a regression model. Assume there are T weak segmenters and the outputs of them are binary images $b_t(x)$, $t = 1, 2, \dots, T$ distinguishing the structures from the background. For a given training image $I(x)$, $\Phi_t(x)$ is the signed distance transform computed from the t^{th} weak segmentation output $b_t(x)$ and $L(x)$ is the signed distance transform of the ground truth label image, representing the true segmentation of $I(x)$. The algorithm assumes that $L(x)$ is the locally weighted combination of $\Phi_t(x)$.

Since the parameters to be optimized depend on both the voxel locations and the weak segmenters, we solve this minimization problem voxel-wise by re-arranging each $\Phi_t(x)$ into a column vector and stacking them together column by column in the $M \times T$ matrix $\bar{\Phi}$, where M is the number of voxels in the image. Let the column vector \mathbf{l} be the re-arrangement of $L(x)$ and denote column vector ϕ_i as taken from the i^{th} row of $\bar{\Phi}$, l_i as the i^{th} entry of \mathbf{l} and column vector \mathbf{w}_i as the weights associated with ϕ_i .

It is well-justified to assume that the weights \mathbf{w}_i and \mathbf{w}_j are expected to be similar if the j^{th} voxel is in the neighborhood of the i^{th} voxel. We therefore adopt a regularization term based on the pre-computed matrix G , which captures the similarity of the labels within the neighborhood. We eventually formalize our cost function in the following:

$$\mathbf{w}^* = \arg \min_{\mathbf{w}} \sum_{i=1}^M \|\mathbf{w}_i \cdot \phi_i - l_i\|_2^2 + \lambda \sum_{i,j=1}^M G(i, j) \|\mathbf{w}_i - \mathbf{w}_j\|_2^2.$$

A closed form solution can be derived for this objective function. We start by expanding the cost function and get

$$E(\{\mathbf{w}_i\}) = \sum_{i=1}^M \mathbf{w}_i^t \phi_i \phi_i^t \mathbf{w}_i + \sum_{i=1}^M l_i^2 - 2 \sum_{i=1}^M l_i \mathbf{w}_i^t \phi_i + \lambda \sum_{i,j=1}^M G(i,j)(\mathbf{w}_i^t \mathbf{w}_i + \mathbf{w}_j^t \mathbf{w}_j - 2\mathbf{w}_i^t \mathbf{w}_j).$$

With the following notations: (1) $H_i = \phi_i \phi_i^t$, (2) $W^t = [\mathbf{w}_1^t, \dots, \mathbf{w}_M^t]$, (3) $B_k = H_k + 2\lambda(\sum_{i=k,j \neq 1} G(i,j) + \sum_{i \neq 1,j=k} G(i,j))I_{T \times T}$, (4) $p^t = [l_1 \phi_1^t, \dots, l_M \phi_M^t]$ and after some algebra, the cost function is re-arranged into a matrix form:

$$E = W^t \begin{pmatrix} B_1 & -2\lambda G(1,2)I_{T \times T} & \dots & -2\lambda G(1,N)I_{T \times T} \\ -2\lambda G(2,1)I_{T \times T} & B_2 & \dots & -2\lambda G(2,N)I_{T \times T} \\ \dots & \dots & \dots & \dots \\ -2\lambda G(N,1)I_{T \times T} & -2\lambda G(N,2)I_{T \times T} & \dots & B_N \end{pmatrix} W - 2p^t W + \sum_{n=1}^N l_n^2.$$

We take the derivative of E w.r.t. W and set the result to 0 in order to solve for the weights. We then have $\frac{\partial E}{\partial W} = (D^t + D)W - 2p^t = 0$, with D being the matrix in the equation above that contains B_k as diagonal. The problem is finally reduced to solving the following linear system $(D^t + D)W = 2p^t$. Note that since $D^t + D$ is a sparse matrix, we finally solve a *sparse least-squares problem* which can be efficiently performed.

2.3 Testing Stage: Sparse Linear Combination

Assume there are N training images and for each image, we solve for the local weights W to combine the weak segmenters. We denote by W_{nt} , the weight matrix for the n^{th} training image and the t^{th} weak segmenter, which is basically a single matrix involving the re-arrangement of W that we solved for in the training stage. To avoid overfitting, in the testing stage, not all the training results are used. *This is similar to the situation when a new patient comes in for medical consultation, we expect that a good strategy for the experts involves searching for useful relevant case studies from the old patients in order to arrive at a consensus diagnosis.* Therefore, only a subset of the trained parameters are helpful in testing. Several techniques can be used to achieve this goal, for instance the K-Nearest Neighbor (kNN) and the Sparse Representation methods. The kNN based search for the most similar cases to represent the testing data will potentially fail when all the training images differ from the test sample. Besides, one has to resort to a relatively complicated data structure for fast kNN implementation when the feature dimension is high. Recently, sparsity has been

investigated for feature selection [10]. Following this philosophy, we compute the optimal sparse combination of the training images to approximate the test data.

Formally, let A be a matrix with N columns where the i^{th} column contains the i^{th} training image and let Y be the given test image. To obtain a sparse combination of training set images, we use an L_1 norm regularizer. The problem is formalized as follows: $\gamma^* = \arg \min_{\gamma} \|A\gamma - Y\|_2^2 + \alpha \|\gamma\|_1$, which can be solved using existing techniques, such as LASSO. The final strong segmenter is then given by $S(x) = \sum_{n=1}^N \sum_{t=1}^T \gamma_n W_{nt}(x) \Phi_t(x)$.

3 Experiments

In this section, we empirically validate our SegMix algorithm and compare it with the widely used voting based methods. Our algorithm significantly improves the weak segmentation results given a small number of low-accuracy weak segmenters.

Hippocampus Data Set: This experiment is performed on a hippocampal data set containing 60 brain MRI (T1) images, with the right hippocampi manually segmented by an expert neurologist. We divide the data set into 2 groups. The first group contains 20 images used in building the multiple atlases. 10-fold cross validation is applied to the remaining 40 images. The original brain images are first corrected for intensity inhomogeneity and normalized, then registered to the same coordinate system using a similarity transformation. Since the hippocampi are within a certain region of the brain, we therefore define a bounding box that approximately encloses each hippocampus and only take these ROIs as the input to our segmentation algorithm. The size of ROI is $56 \times 39 \times 30$. We extract the ROI for each test image by first deforming the brain MRI scan to a labeled brain template and finding the ROI based on the template information.

Weak Segmenters: *Note that any segmentation method is applicable as a weak segmenter within our framework.* However, in order to demonstrate the robustness and performance of our algorithm and compare it to the existing multi-atlas segmentation methods, we use atlas-based segmentation [1] as the weak segmenter. We first cluster the 20 images into a set of groups based on the hippocampal shape information from the labels. The signed distance transform representation of each shape is mapped to the square-root density space via a Schrödinger wave function [8]. Hence, each shape corresponds to a single point on the high dimensional sphere and the similarity between the shapes is computed intrinsically using the geodesic distance on the unit sphere. Armed with this intrinsic similarity measure, any clustering method may be used here. We employ affinity propagation [11] since it does not require the number of clusters to be specified. We get 5 clusters and the atlases/centers for each group are used for the weak segmentations.

Performance Measure: The evaluation metrics for measuring the performance of the algorithm used in this paper include the *similarity index* (known as Dice coefficient) $SIM = \frac{2V(A \cap B)}{V(A) + V(B)}$, which computes the overlap of two volumes and

the *difference index* $DIF = \frac{2|\mathcal{V}(A) - \mathcal{V}(B)|}{\mathcal{V}(A) + \mathcal{V}(B)}$, which measures the size difference of the two volumes. Here $\mathcal{V}(\cdot)$ computes the volume. A good segmentation has larger overlap with the ground truth hence higher similarity index, but lower difference index.

Experiment Setting: The free parameters involved in our SegMix algorithm include λ and α for the regularization in the training and testing stages respectively. We (empirically) discovered that the algorithm is not very sensitive to the particular choice of λ and α , so we set $\lambda = 0.5$ and $\alpha = 0.5$ throughout the comparison experiments. An 18-connected neighborhood is used for computing the graph G . The nonrigid registration algorithm used for atlas-based segmentation is Demons, where we use all the *default* parameter settings, i.e. all the registrations are performed with the same parameters. *Due to the use of standard defaults, atlas-based segmentation leads to low-accuracy for each weak segmenter.* However, the experimental results indicate that SegMix significantly improves on those weak segmentations.

Experimental Results: To validate our proposed algorithm, the experiments are performed on the 3D hippocampal MRI images using (1) SegMix, (2) Global Weighted Voting (GWV) and (3) Local Weighted Voting (LWV) [4] with 5 weak segmenters. In the following table, we list the average performance evaluation for the 10-fold cross validation of the 40 images.

Table 1. The average SIM and DIF indices for SegMix, GWV and LWV

	weak SIM	final SIM	Increased	weak DIF	final DIF	Decreased
SegMix	0.68 ± 0.04	0.80 ± 0.03	17.65%	0.42 ± 0.10	0.09 ± 0.06	78.57%
GWV	0.68 ± 0.04	0.73 ± 0.04	7.35%	0.42 ± 0.10	0.37 ± 0.11	11.90%
LWV	0.68 ± 0.04	0.74 ± 0.04	8.82%	0.42 ± 0.10	0.39 ± 0.10	7.14%

Since better segmentation corresponds to larger SIM but smaller DIF, we show the increased SIM value and decreased DIF value w.r.t. the weak segmentations. Due to the low accuracy and limited number of the weak segmenters, the performance of the voting-based methods are poor as expected. We also present the SIM and the DIF values for both the weak segmentations and the

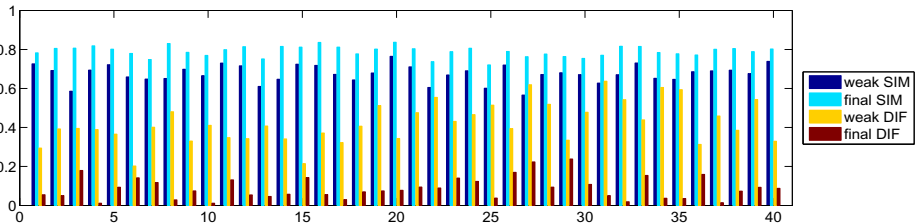


Fig. 2. The figure shows the SIM and DIF for each image

final strong segmentation and show the improvements of our techniques w.r.t. the weak segmenters for each test image in Fig 2.

4 Conclusion

In this paper, we introduced a novel weak segmentation combination strategy based on the assumption that the locally weighted combination varies w.r.t. *both* the weak segmenters and the training images. We learned the weighted combination during the training stage using a discriminative spatial regularization which depends on training set labels. In the testing stage, a sparse regularization scheme was imposed to avoid overfitting. The experimental results indicated that our algorithm not only outperforms the voting based methods but also significantly improves the performances of the weak segmenters.

References

1. Carmichael, O.T., Aizenstein, H.A., Davis, S.W., Becker, J.T., Thompson, P.M., Meltzer, C.C., Liu, Y.: Atlas-based Hippocampus Segmentation in Alzheimer's Disease and Mild Cognitive Impairment. *NeuroImage* 27(4), 979–990 (2005)
2. Golland, P., Grimson, W., Shenton, M., Kikinis, R.: Detection and Analysis of Statistical Differences in Anatomical Shape. *Med. Image Anal.* 9(1), 69–86 (2005)
3. Morra, J.H., Tu, Z., Apostolova, L.G., Green, A., Toga, A.W., Thompson, P.M.: Comparison of AdaBoost and Support Vector Machines for Detecting Alzheimer's Disease through Automated Hippocampal Segmentation. *IEEE Trans. Med. Image* 29(1), 30–42 (2010)
4. Artaechevarria, X., Munoz-Barrutia, A., Ortiz-de-Solorzano, C.: Combination Strategies in Multi-atlas Image Segmentation: Application to Brain MR Data. *IEEE Trans. Med. Image* 28(8), 1266–1277 (2009)
5. Khan, A., Cherbuin, N., Wen, W., Anstey, K.J., Sachdev, P., Beg, M.F.: Optimal Weights for Local Multi-atlas Fusion using Supervised Learning and Dynamic Information (SuperDyn): Validation on Hippocampus Segmentation. *NeuroImage* 56(1), 126–139 (2011)
6. Kumar, S., Hebert, M.: Discriminative Random Fields. *Int. J. Comput. Vision* 68(2), 179–201 (2006)
7. Pohl, K.M., Fisher, J., Shenton, M., McCarley, R.W., Grimson, W.E., Kikinis, R., Wells, W.M.: Logarithm Odds Maps for Shape Representation. In: Larsen, R., Nielsen, M., Sparring, J. (eds.) *MICCAI 2006, Part II*. LNCS, vol. 4191, pp. 955–963. Springer, Heidelberg (2006)
8. Chen, T., Rangarajan, A., Eizenschenk, S.J., Vemuri, B.C.: Construction of Neuroanatomical Shape Complex Atlas from 3D Brain MRI. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) *MICCAI 2010, Part III*. LNCS, vol. 6363, pp. 65–72. Springer, Heidelberg (2010)
9. Leventon, M.E., Grimson, W.E.L., Faugeras, O.: Statistical Shape Influence in Geodesic Active Contours. In: *IEEE Conf. CVPR*, pp. 316–323 (2000)
10. Zhang, S., Huang, J., Huang, Y., Yu, Y., Li, H., Metaxas, D.: Automatic Image Annotation Using Group Sparsity. In: *IEEE Conf. CVPR*, pp. 3312–3319 (2010)
11. Frey, B.J., Dueck, D.: Clustering by Passing Messages Between Data Points. *Science* 315, 972–976 (2007)

Active Learning for Interactive 3D Image Segmentation

Andrew Top¹, Ghassan Hamarneh¹, and Rafeef Abugharbieh²

¹ Medical Image Analysis Lab, Simon Fraser University

² Biomedical Signal and Image Computing Lab, University of British Columbia
{atop,hamarneh}@sfu.ca, rafeef@ece.ubc.ca

Abstract. We propose a novel method for applying active learning strategies to interactive 3D image segmentation. Active learning has been recently introduced to the field of image segmentation. However, so far discussions have focused on 2D images only. Here, we frame interactive 3D image segmentation as a classification problem and incorporate active learning in order to alleviate the user from choosing where to provide interactive input. Specifically, we evaluate a given segmentation by constructing an “uncertainty field” over the image domain based on boundary, regional, smoothness and entropy terms. We then calculate and highlight the plane of maximal uncertainty in a batch query step. The user can proceed to guide the labeling of the data on the query plane, hence actively providing additional training data where the classifier has the least confidence. We validate our method against random plane selection showing an average DSC improvement of 10% in the first five plane suggestions (batch queries). Furthermore, our user study shows that our method saves the user 64% of their time, on average.

1 Introduction

3D image segmentation is one of, if not the most, important and ubiquitous tasks in medical image analysis. Fully-manual slice by slice segmentation of images is widely recognized as infeasible, being too tedious, time consuming, expensive, and suffering from high inter- and intra-operator variability. Furthermore, manual operators have very limited ability to integrate the information available in 3D data given that almost all displays and data entry mechanisms are 2D. On the other side of the spectrum, fully-automated segmentation techniques have generally struggled to achieve the accuracy and robustness levels needed for clinical practice. Consequently, highly automated interactive image segmentation approaches [8] have recently become the approach of choice in most real life medical applications, and thus are garnering the focus of the medical image analysis community.

Many interaction mechanisms have been previously pursued such as interactive contour delineation [7] or region seeding [3]. Interactive segmentation is often an iterative process. A user provides input that guides computations which return output back to the user so that they may provide additional input. Ideally, the interaction process can occur in real-time so that users receive immediate feedback on their actions. In the case where user interaction takes the form of specifying which voxels belong to the object or region of interest (ROI) and which do not, the interactive segmentation process can be nicely formulated within a supervised machine learning framework. The user input

can thus be seen as labeled training data, which the interactive segmentation algorithm, or classifier, uses to label the remaining data (i.e. to segment the ROI).

Active learning (AL) [11] refers to supervised machine learning where, instead of the user, a component of the algorithm called the ‘query strategy’ is responsible for choosing the training data to be labeled. The query strategy, having intimate knowledge of the inner workings of the specific classification algorithm deployed, is in an excellent position to determine candidate high-quality training data voxels. By labeling voxels according to AL queries, interactive segmentation algorithms can produce improved segmentations using less, more efficient user input.

Image segmentation has had a number of treatments in the literature that relate to AL. Ma et al. [6] explored image segmentation under an AL framework where they used a support vector machine as the segmentation method and uncertainty sampling as the query strategy. Pavlopoulou et al. [9] discussed the translation of interactive contour delineation to an AL formulation. They used the term “active learning” in reference to the algorithm’s support for interactive correction of the contour should it deviate from the ground truth, but this definition differs from that of formal AL. Li et al. [5] explored the segmentation of hyperspectral images by employing AL to query the user for more seedpoints when confidence is lacking. Queries were made in the form of a set of 2D image pixels labeled with low confidence, but it is unclear how to present these queries in a user-friendly fashion. In [12], we presented a new approach to optimize user interaction using a ‘Spotlight’ feature that was recently implemented into the software package TurtleSeg¹. That work however was not formulated in an AL framework.

None of the previous methods treated AL in the context of 3D segmentation, which is the focus of our work in this paper. We solve two challenges: 1. The 3D image must be presented to the user on a 2D display in a way that best reveals the object. Once constructed (using some method), a candidate segmentation should be properly overlaid in context. 2. The user must assess the entire 3D segmentation in order to improve it by providing additional input. The simple inspection of the entire image and its segmentation, which is straightforward in 2D, becomes a much more complicated issue in 3D. In order to deal with the first challenge, we adopt the multi-planar reconstruction (MPR) technique where orthogonal or oblique slices of the 3D image are presented to the user. MPR is preferred as it is the most common 3D image display technique in clinical practice. To address the second challenge, we employ an AL uncertainty sampling query strategy, which is designed to highlight to the user the regions in 3D where additional training data will likely improve the segmentation. To the best of our knowledge, we are the first to formulate 3D interactive image segmentation in an AL framework.

2 Methods

For clarity of exposition, we start by defining our notation. Let $\Omega \subset \mathbb{R}^3$ represent the spatial image domain, $y : \Omega \rightarrow \{0, 1\}$ represent the classifier labeling, $I : \Omega \rightarrow \mathbb{R}$ represent the image intensity function, and $U : \Omega \rightarrow \mathbb{R}$ represent the classification uncertainty field. We formulate our 3D image segmentation as a data classification problem. Our feature vector consists of pixel locations $\mathbf{x} \in \Omega$ and intensity $I(\mathbf{x})$,

¹ <http://www.turtleseg.org>

but since the intensity depends on the position, we hereafter refer to the feature vector as \mathbf{x} for simplicity. The solution to the classification problem is to determine a label $y(\mathbf{x}) \in \{0, 1\}$ for each $\mathbf{x} \in \Omega$. We adopt a supervised learning approach for interactive segmentation. The user provides a set of training data, T , in which each element has the form $(X, Y) \in T$, where $X \in \Omega$, $Y \in \{0, 1\}$ and each X is unique, that is $\forall (X_a, Y_a), (X_b, Y_b) \in T, X_a = X_b \implies a = b$. The training data is used to train a classifier and obtain an image of labels i.e. a segmentation, y . The classifier can be any probabilistic or energy minimizing 3D image segmentation algorithm, although probabilistic algorithms (for example [3]) are preferred since they naturally encode uncertainty in the results. Learning occurs as the training data is augmented, improving the ability of the classifier to label unlabeled image pixels.

Since the initial set of training data is typically not sufficient for the classifier to produce accurate labelings, we assess the uncertainty, U , of the classification results. Should the uncertainty remain above a preset threshold or the user deems the segmentation to be of insufficient quality, an AL batch query is calculated from U in order to present to the user regions of maximal uncertainty. Specifically, we find the plane that passes through the most uncertain regions of the classification. The user's resulting input augments the current training data, and the learned classifier is re-executed.

2.1 Uncertainty Field of Automatic Segmentation

The uncertainty value at a point \mathbf{x} in the image, $U(\mathbf{x})$, reflects the lack of confidence of the classifier in its classification $y(\mathbf{x})$ at point \mathbf{x} . We propose a novel composite uncertainty field U , comprising a weighted sum of four sub fields.

$$U(\mathbf{x}, y) = \lambda_E U_E(\mathbf{x}, p_1(\mathbf{x})) + \lambda_B U_B(\mathbf{x}, y) + \lambda_R U_R(\mathbf{x}, y) + \lambda_S U_S(\mathbf{x}, y), \quad (1)$$

where U_E is an entropy energy that depends on the probability $p_1 : \Omega \rightarrow [0, 1]$ of point \mathbf{x} being labeled 1. U_B is a boundary energy, U_R is a regional energy, and U_S is a smoothness energy. Note that we are not interested in finding a classification $y(\mathbf{x})$ that minimizes the sum in (1) over the image domain, but rather how the value of U at one position in the image compares to the value of U at another. For example, a classifier which minimizes the boundary term U_B given a set of constraints might still give a solution with high U_B in regions where there are no hard edges to follow. When using a probabilistic segmentation algorithm as a classifier, a natural choice for assessing uncertainty is the entropy of the segmentation results, giving the entropy energy as

$$U_E(\mathbf{x}, p_1(\mathbf{x})) = -p_1(\mathbf{x}) \log_2 p_1(\mathbf{x}) - (1 - p_1(\mathbf{x})) \log_2 (1 - p_1(\mathbf{x})). \quad (2)$$

For the boundary energy, we adopt from [1] a function of the image intensity gradient,

$$U_B(\mathbf{x}, y) = \delta(D_s(\mathbf{x}, y)) \frac{1}{1 + |\nabla I(\mathbf{x})|^\alpha}, \quad (3)$$

where $D_s(\mathbf{x}, y)$ gives the distance from \mathbf{x} to the classification boundary that divides points where $y(\mathbf{x}) = 0$ and $y(\mathbf{x}) = 1$. δ is the delta function which allows us to consider the gradient only for points on the segmentation boundary. In this paper, we set $\alpha = 2$.

This term will have values near 0 along a segmentation surface that follows a strong edge in the image data, and values near 1 in homogeneous areas. See Section 3.1 for details on how $\delta(D_s(\mathbf{x}, y))$ is implemented. We define the regional energy term as in [2] to measure how well each voxel inside and outside the proposed segmentation conforms to a respective intensity distribution. We derive the intensity distributions for foreground and background voxels by calculating the maximum likelihood Gaussian distribution parameters on the training data. For Gaussian foreground and background intensity distributions given respectively as $p(I(\mathbf{x})|Y = 1) = \mathcal{N}(\mu_1, \sigma_1)$ and $p(I(\mathbf{x})|Y = 0) = \mathcal{N}(\mu_0, \sigma_0)$, we can write the regional energy using Baye’s theorem as

$$U_R(\mathbf{x}, y) = p(Y = y(\mathbf{x})|I(\mathbf{x})) = \frac{p(I(\mathbf{x})|Y = y(\mathbf{x}))}{p(I(\mathbf{x})|Y = 0) + p(I(\mathbf{x})|Y = 1)} \tag{4}$$

where we assume $p(Y = 0) = p(Y = 1) = 0.5$. Finally, since smooth surfaces have less area, we define a smoothness energy term based on the surface area of the segmentation around a point \mathbf{x} . With $N_{\mathbf{x}}$ representing a local 3D neighborhood around \mathbf{x} , we have

$$U_S(\mathbf{x}, y) = \iiint_{N_{\mathbf{x}}} \delta(D_s(\mathbf{z}, y)) \, dV. \tag{5}$$

2.2 Batch Query Active Learning Based on Planar Slices

Given an existing classification, heuristically the user should provide labels in the most uncertain regions in order to best improve the classification. Point by point labeling is too slow, so we instead use MPR to present the user with a plane of uncertain points so that they can be labeled simultaneously. Querying the user for multiple labels at the same time, as we do here, is known as a batch query [11]. Therefore, we seek the plane with maximal uncertainty, which we define as $\operatorname{argmax}_{\mathcal{P}}(U_{\mathcal{P}})$ where $U_{\mathcal{P}}$ is given as

$$U_{\mathcal{P}} = \iint_{\mathcal{P}} U(\mathbf{x}) \, dA = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} U(f_{\mathcal{P}}(u, v)) \left| \frac{\partial f_{\mathcal{P}}}{\partial u} \times \frac{\partial f_{\mathcal{P}}}{\partial v} \right| \, dudv. \tag{6}$$

Here, $f_{\mathcal{P}}$ is a function mapping \mathbb{R}^2 to \mathcal{P} , and $\left| \frac{\partial f_{\mathcal{P}}}{\partial u} \times \frac{\partial f_{\mathcal{P}}}{\partial v} \right|$ is introduced with the change of variables. We evaluate (6) by calculating the uncertainty at sample points on \mathcal{P} . Since in general the uncertainty field may be an arbitrary scalar image, (6) may have many local maxima with respect to \mathcal{P} . We therefore run multiple iterations of gradient descent with randomized initial parameters (Section 3.1) in order to determine an approximate solution to $\operatorname{argmax}_{\mathcal{P}} U_{\mathcal{P}}$.

To that end, we first parameterize \mathcal{P} by its normal $\mathbf{n}_{\mathcal{P}}$, and a point on \mathcal{P} , $\mathbf{p}_{\mathcal{P}}$. Let $f_{\mathcal{P}}(u, v) = \mathbf{p}_{\mathcal{P}} + u\mathbf{a} + v\mathbf{b}$, where \mathbf{a} and \mathbf{b} are orthonormal vectors perpendicular to $\mathbf{n}_{\mathcal{P}}$, then $\left| \frac{\partial f_{\mathcal{P}}}{\partial u} \times \frac{\partial f_{\mathcal{P}}}{\partial v} \right| = 1$. Making use of the chain rule, we calculate $\nabla_{\mathbf{n}_{\mathcal{P}}} U_{\mathcal{P}}$ as

$$\nabla_{\mathbf{n}_{\mathcal{P}}} U_{\mathcal{P}} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (uJ_{\mathbf{a}, \mathbf{n}_{\mathcal{P}}}^T + vJ_{\mathbf{b}, \mathbf{n}_{\mathcal{P}}}^T) \nabla_x U(f_{\mathcal{P}}(u, v)) \, dudv, \tag{7}$$

where $\nabla_x U(f_{\mathcal{P}}(u, v))$ is the gradient of the uncertainty field with respect to the spatial coordinates x . We use the notation $J_{t,s}$ to represent the Jacobian of t with respect to s . Using the chain rule again, $\nabla_{\mathcal{P}} U_{\mathcal{P}}$ is easily shown to be

$$\nabla_{\mathcal{P}} U_{\mathcal{P}} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \nabla_x U(f_{\mathcal{P}}(u, v)) \, du \, dv \quad (8)$$

since $J_{f_{\mathcal{P}}, \mathcal{P}} = I$. The maximal plane is presented to the user who will then classify it. This 2D segmentation is easy to do and well-studied in computer vision. There are many methods to choose from: [7][10][3] and others. All these can directly or indirectly provide a labeling of the plane.

3 Results

We experimentally verify our automatic plane suggestion approach by comparing it to randomly chosen, as well as user-chosen, slice planes. Since our approach is general enough to be applicable to a number of classifiers, we first describe our specific implementation in Section 3.1 before presenting qualitative and quantitative results in Section 3.2 and Section 3.3 respectively.

3.1 Implementation Details

We use the Random Walker (RW) [3] algorithm as our classifier (i.e. assigns a class label to each pixel), since it is capable of producing a probabilistic segmentation, it accepts as input a set of pre-labeled feature vectors, and it naturally applies to 3D images. To allow RW to run in a feasible amount of time on 3D images, we use techniques presented in [4] to develop a GPU RW implementation using NVIDIA’s CUDA API. In order to keep the classifier consistent for our tests, we set the RW algorithm’s only parameter to $\beta = 10$, a value that was found to produce good segmentations. The $D_s(x, y)$ distance values seen in (3) and (5) are implemented using distance maps. The $\delta(x)$ functions are approximated by $e^{-\frac{x^2}{2\sigma}}$, with $\sigma = 1$, giving a differentiable “soft delta” function. In our experimental software, users provide training data along slices using 2D Livewire, which are rasterized to seedpoints in the 3D image and then passed to RW as input. At any time after a segmentation has been computed the user can click a button to have a batch query slice presented to them. For the following studies, we set $\lambda_E = 0.80$, $\lambda_B = 0.05$, $\lambda_R = 0.15$ and $\lambda_S = 0$, which we found empirically to give good results and were justified because U_E exactly captures entropy while others less so. The plane that maximizes (6) is found by initializing a plane 36 times to a random orientation and ensuring the plane always passes through the the existing segmentation. For each plane initialization, gradient descent is iterated 250 times, and the optimal plane over the 36 runs is returned as the query plane.

3.2 Qualitative Results

We evaluate our method qualitatively by observing that it makes intelligent decisions for each slice suggestion. We validate by segmenting two different medical images in

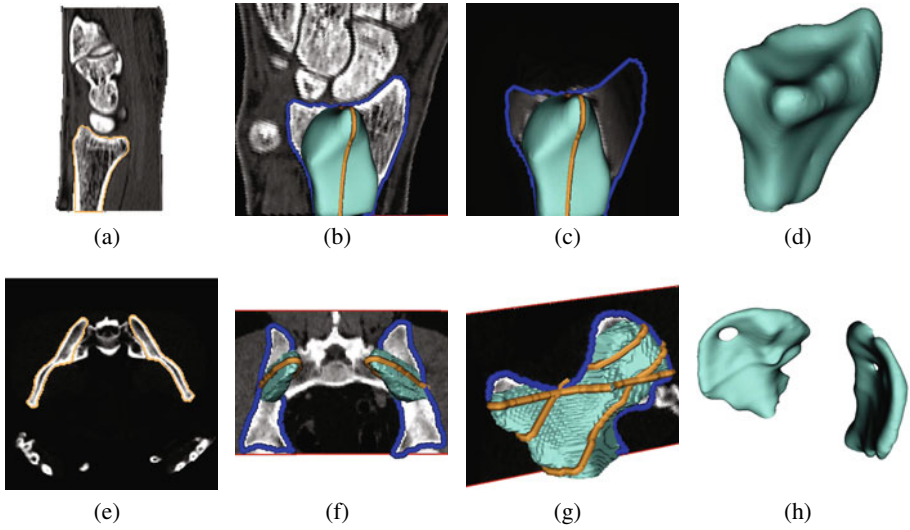


Fig. 1. Example snapshots during the AL interactive segmentation process. The first row shows the radius bone in a CT image. (a) A slice of the image, and a single initial contour in orange. (b) The results of the initial segmentation surface, the next AL suggested plane, and its user-provided classification (blue). (c) A slice through the uncertainty field. (d) The final 3D segmentation after 3 more active queries were labeled. The second row shows the segmentation of the iliac bones in a pelvis CT image. (e) A slice through the intensity image as well as the initial contours. (f) The first AL query slice. (g) The 3rd AL query slice. (h) The final segmentation after 12 slices have been classified.

Figure 1 demonstrating the active learning queries. In the first row, the radius bone in a wrist CT image is segmented. Figure 1(b) shows the first active query slice, along with its segmentation (blue). Note that the slice passes through a large portion of the radius bone that was, in error, not previously included in the initial segmentation. Looking at the uncertainty field through the same slice in figure 1(c), we see that high uncertainty is detected in the radius bone region not included in the initial segmentation. In the second row, the iliac bones of a pelvis CT image are segmented. We see in figure 1(f) that the query slice is targeting a large portion of the iliac bones that are contained in the initial segmentation. Figure 1(g) shows the third AL query slice beginning to focus on the left iliac bone, in order to better finalize that region of the segmentation.

3.3 Quantitative Validation

We test our method on eight different datasets: radius bone in a CT, femur bone in an MRI, liver in a CT, two images of a vastus lateralis muscle in an MRI, putamen in an MRI, tibia bone in an MRI, and vastus intermedius muscle in an MRI. Image sizes ranged from 150^3 to 190^3 pixels where suggestions took around 20 to 33 seconds, respectively. We report accuracy increase after each training data augmentation step. We first compare our proposed active query planes with random plane selection. Random plane selection involves randomly choosing a plane orientation and offset, constrained

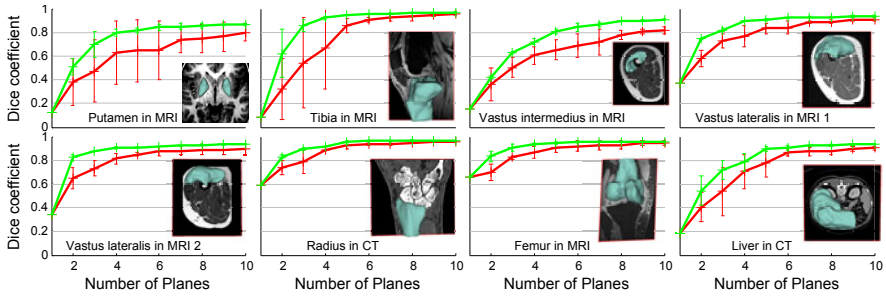


Fig. 2. A comparison of our AL technique to random plane selection. Each graph shows the DSC value of a segmentation versus the number of slices classified. The green curves show the results of using AL for plane selection, while the red curves show the result of randomly choosing planes. Sample slices of each segmented 3D volume are shown to clarify the image example context.

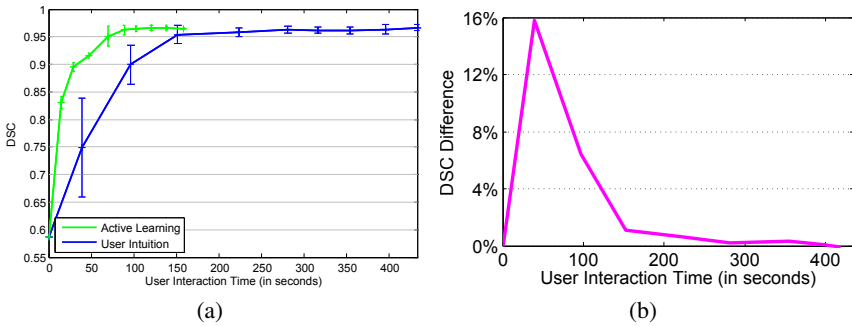


Fig. 3. Results of the user study on the “Radius in CT” image. (a) An errorbar plot is given showing our AL approach improving the segmentation faster than user decisions. (b) The average difference in DSC percentage versus time.

to pass through the existing segmentation. Since both methods are non-deterministic, we re-segment each dataset three times and average the results. The accuracy is shown in terms of the Dice Similarity Coefficient (DSC) between the produced segmentation and the ground truth segmentation. For all results, in order to maintain consistency, we automatically derive the user segmentations over slice planes by extracting them from the ground truth segmentation. Figure 2 shows the comparison results between AL queries and random plane selection. Notice that the AL queries lead to higher DSC maxima, the DSC increases faster per user-segmented slice, and the average DSC increases by 10% on the first five plane suggestions (batch queries). Often, the first active query suggestion results in a better segmentation than multiple random slice planes.

Next we present the results of a 4 person user-study comparing our method of plane selection to human intuition on “Radius in CT” dataset chosen from above. Figure 3 gives a plot for the dataset showing DSC increase versus user interaction time where errorbars show the variance of the different users’ output. Results clearly demonstrate how our AL approach significantly reduces the time needed to achieve any DSC value. Specifically, a total average reduction of 64% in user interaction time was achieved.

4 Conclusions

In this paper we formulate interactive 3D image segmentation in an AL framework. Our contributions include being the first to formulate interactive 3D image segmentation as a formal AL process. In doing so, we alleviate the user from choosing where to provide interactive input. In order to facilitate the batch queries, we automate assessment of uncertainty in the classification as well as the search through the uncertainty for the best query plane. Finally, we show our method to significantly reduce the required user input in interactive 3D image segmentation tasks through a user study. Future work includes incorporating this technique into other interactive segmentation algorithms and learning the energy functional weights automatically via training data.

References

1. Caselles, V., Kimmel, R., Sapiro, G.: Geodesic active contours. *Int. J. Comput. Vision* 22(1), 61–79 (1997)
2. Chan, T., Vese, L.: Active contours without edges. *IEEE Trans. Image Proc.* 10(2), 266–277 (2001)
3. Grady, L.: Random walks for image segmentation. *IEEE Trans. Pat. Anal. and Mach. Intel.* 28(11), 1768–1783 (2006)
4. Grady, L., Schiwietz, T., Aharon, S., Westermann, R.: Random walks for interactive organ segmentation in two and three dimensions: Implementation and validation. In: Duncan, J., Gerig, G. (eds.) *MICCAI 2005. LNCS*, vol. 3750, pp. 773–780. Springer, Heidelberg (2005)
5. Li, J., Bioucas-Dias, J., Plaza, A.: Supervised hyperspectral image segmentation using active learning. In: *2010 2nd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*, pp. 1–4 (2010)
6. Ma, A., Patel, N., Li, M., Sethi, I.: Confidence based active learning for whole object image segmentation. In: Günsel, B., Jain, A., Tekalp, A., Sankur, B. (eds.) *MRCSS 2006. LNCS*, vol. 4105, pp. 753–760. Springer, Heidelberg (2006)
7. Mortensen, E.N., Barrett, W.A.: Intelligent scissors for image composition. In: *22nd Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH*, pp. 191–198. ACM, New York (1995)
8. Olabarriaga, S., Smeulders, A.: Interaction in the segmentation of medical images: a survey. *Med. Image Anal.* 5(2), 127–142 (2001)
9. Pavlopoulou, C., Kak, A.C., Brodley, C.: Application of semi-supervised and active learning to interactive contour delineation. In: *ICML 2003 Workshop on the Continuum from Labeled to Unlabeled Data in Machine Learning and Data Mining*, pp. 26–33 (2003)
10. Rother, C., Kolmogorov, V., Blake, A.: *GrabCutTM*: Interactive foreground extraction using iterated graph cuts. *ACM Transactions on Graphics* 23, 309–314 (2004)
11. Settles, B.: *Active learning literature survey*. Tech. Rep. 1648, University of Wisconsin-Madison (2010)
12. Top, A., Hamarneh, G., Abugharbieh, R.: Spotlight: Automated confidence-based user guidance for increasing efficiency in interactive 3D image segmentation. In: Menze, B., Langs, G., Tu, Z., Criminisi, A. (eds.) *MICCAI 2010. LNCS*, vol. 6533, pp. 204–213. Springer, Heidelberg (2011)

3D Segmentation of Rodent Brain Structures Using Hierarchical Shape Priors and Deformable Models

Shaoting Zhang¹, Junzhou Huang¹, Mustafa Uzunbas¹, Tian Shen²,
Foteini Delis³, Xiaolei Huang², Nora Volkow³, Panayotis Thanos³,
and Dimitris N. Metaxas¹

¹ CBIM, Rutgers, The State University of New Jersey, Piscataway, NJ, USA

² Computer Science and Engineering Department, Lehigh University, PA, USA

³ Brookhaven National Lab, NY, USA

Abstract. In this paper, we propose a method to segment multiple rodent brain structures simultaneously. This method combines deformable models and hierarchical shape priors within one framework. The deformation module employs both gradient and appearance information to generate image forces to deform the shape. The shape prior module uses Principal Component Analysis to hierarchically model the multiple structures at both global and local levels. At the global level, the statistics of relative positions among different structures are modeled. At the local level, the shape statistics within each structure is learned from training samples. Our segmentation method adaptively employs both priors to constrain the intermediate deformation result. This prior constraint improves the robustness of the model and benefits the segmentation accuracy. Another merit of our prior module is that the size of the training data can be small, because the shape prior module models each structure individually and combines them using global statistics. This scheme can preserve shape details better than directly applying PCA on all structures. We use this method to segment rodent brain structures, such as the cerebellum, the left and right striatum, and the left and right hippocampus. The experiments show that our method works effectively and this hierarchical prior improves the segmentation performance.

1 Introduction

Magnetic resonance imaging (MRI) at a spatial resolution of at least 100 μ m in one dimension is frequently referred to as MR microscopy (MRM) and is currently available with the use of high magnetic field images. This technical achievement has permitted the detailed anatomical study of the rodent brain, which is much smaller than the human brain and requires a small voxel size in order to be imaged. Rodents are often used as models of human disease not only because they frequently exhibit key features of abnormal neurological conditions but also because they are a convenient starting point for novel studies. The analysis of rodent brain image faces similar challenges to human imaging, with

individual variation in size, morphology, and topology of the brain structures complicating the neuroanatomical studies. Such analysis is frequently performed by segmenting the regions of interest (ROI) in rodent brain images. The challenge of this segmentation task is threefold. 1) The image information is sometimes incomplete or misleading. For example, there is no obvious boundary in part of the striatum, and the cerebellum contains interleaving texture (Fig. 1), 2) It is important to discover and/or preserve some complex local shape details, such as the paraflocculi (i.e., two protruding features of the cerebellum), 3) Given limited number of training samples, it is desirable to segment all interested structures simultaneously by effectively learning a shape atlas for multiple structures.

In recent decades, many deformable model based segmentation methods have been proposed to solve these challenges and achieved tremendous success [6,2,7,11]. The traditional Snakes [6] solely rely on the image gradient information, so they may be trapped by noise and spurious edges. Region analysis strategies [13] have been incorporated in Snake-like models to improve their robustness to noise. Metamorphs [5] was proposed to be able to integrate shape and appearance in a unified space. The model has not only boundary shape but also interior appearance, making it more robust to am-

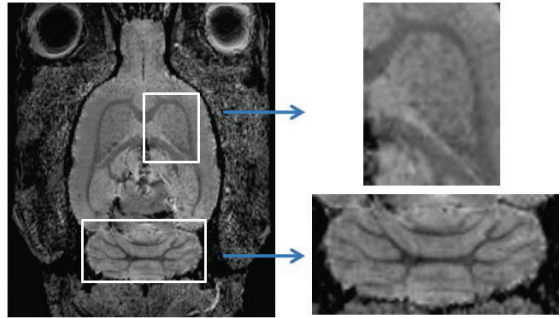


Fig. 1. The MR image of a rat brain used in experiments. The image information is misleading because 1) part of the boundary between striatum (top) and other structures is blurred; 2) the cerebellum (bottom) has complex textures and large gradient values inside.

biguous boundaries and complex internal textures. Its 3D version, Active Volume Model (AVM) [10], is proposed to perform volume segmentation. The AVM model's shape is represented by a simplex mesh and its volumetric interior carries the various visual appearance feature statistics. However, the only shape prior used in AVM is the smoothness constraint. Thus it may not be able to preserve or discover small shape details. Another group of deformable models also considering region information is level set based methods [7,8]. These approaches have been widely used in tubular structure and 3D cortex segmentation tasks since they are topologically free and can be easily used in any dimension. Statistical modeling approaches such as Active Shape Model (ASM) [2] are also widely used. These methods constrain the intermediate shape by using the shape pattern from existing data. Thus they are able to recover or preserve local shape details. However, these methods may need a large amount of 3D training data, whose creation and maintenance can be difficult and time consuming in practice.

Combining AVM and ASM is able to handle both complex texture and shape details without a large set of 3D training data [12]. However, this approach still can not handle multiple structures simultaneously because, with multiple structures together, there is a larger variability and thus it usually requires more training samples in order to capture such variability.

In this paper, we propose a unified algorithm to alleviate these three challenges mentioned above. It includes a robust deformation model which uses both gradient and appearance information, and a shape prior model which can learn shape statistics from a small number of training data by employing a hierarchical modeling scheme. Note that the idea of hierarchical shape prior has been investigated much in the literature [3,4,1]. Our contribution is to combine such idea with a robust deformable model to effectively segment multiple structures of rodent brains. First, the shape statistics of each structure is obtained using topology-aware shape registration and Principal Component Analysis (PCA). The relative positions among structures are also obtained as high level statistics. This hierarchical prior module is able to effectively build shape statistics in two levels even with limited number of training data. Second, the deformation module is used to drive the 3D mesh based on the image information. Its intermediate result is refined and constrained by adaptively using the hierarchical shape priors. Our deformation module is relatively robust to image noise. However, in the presence of bad initialization or heterogeneous texture, the shape prior constraint can still prevent the model from getting stuck or leaking out of the target in the presence of bad initialization or heterogenous texture. We use this algorithm to segment complex structures in rodent brains, such as the cerebellum, the hippocampus and the striatum. Extensive experiments have been designed to evaluate this method.

2 Methodology

Shape Atlas for Multiple Structures: All brain scans are aligned to the same reference brain. Alignment is performed with rigid transformations (rotations and translations), thus after alignment the volume and shape of brain structures do not change. The brain structures are manually segmented by clinical experts. We focus on the cerebellum, the left and right striatum, and the left and right hippocampus. After extracting triangular meshes from the binary image, we use sophisticated geometry processing methods to improve the mesh quality, such as the mesh decimation, isotropic remeshing and detail-preserved smoothing. A topology-aware shape registration method based on Adaptive Focus Deformable Model (AFDM) [9] is then used to register a reference mesh to the others. Thus all processed meshes share the same topology and have one-to-one correspondence among vertices.

Once the training shapes are available, the generalized Procrustes analysis [2] is applied to align them together. PCA is then used to capture the shape statistics, such as the mean shape and its major variances. The shape statistics are captured in a hierarchical scheme. First, we apply PCA on each structure sepa-

rately. Given limited training samples, this approach can learn the shape details better than using all structures. Fig. 2 shows the largest mode of the variations for rodent brain structures. Second, PCA is also used to learn the relative positions among structures, and their locations with respect to the mass centroid. This prior is built at the high level. Both statistics are used as prior information to guide the segmentation process. Since the hierarchical scheme builds priors in two levels, it is able to effectively model the multiple structures effectively even given a small number of training samples.

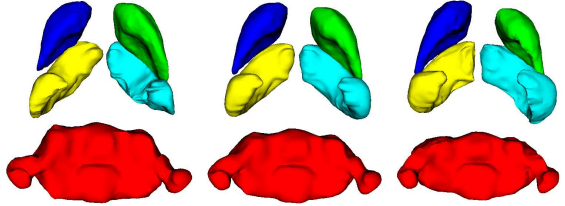


Fig. 2. The largest mode (-3σ to 3σ) of the rat brain variations, which represents the changing of the structure size. Middle is the mean shape.

Deformation Module: Given a test 3D image, it is aligned to the reference brain. Then the mean mass centroid of all structures is used as the estimated center. The mean shape of each structure is placed according to the relative positions learned at the high level statistics. However, this initialization may not be close to the boundary of the testing data because of the variance. Thus deformable models are still needed for accurate segmentation. In order to find the boundary robustly, the initialized models are driven by both gradient and region information derived from the image. Region information alleviates the problems caused by unclear boundaries and complex textures. The overall energy function is: $E = E_{int} + E_{ext} = E_{int} + (E_g + k_R \cdot E_R)$, where E_{int} is the internal (smoothness) energy, E_{ext} is the external (image) energy, E_g is the gradient term, E_R is the region term, k_R is a constant to balance the contributions of the two external energy terms. The balance between the internal and external energies is naturally controlled by the smoothness factor in the stiffness matrix by using Finite Element Method (FEM) as the deformation scheme [10].

The traditional gradient data terms usually include the gradient map, edge distance map, or a combination of both, while the region data term encodes constraints for the model-interior appearance statistics. Considering a module using intensity statistics, the object region is predicted according to the current model-interior intensity distribution. Having both foreground object and background probabilities, we obtain a binary map that represents the predicted object region by applying the Bayesian Decision rule. Connected component analysis is then applied to the binary map to retrieve the connected component that overlaps the current model. This connected region is considered as the current ROI. Let us denote the signed distance transform of the current model's surface shape as Φ_A , and the signed distance transform of the ROI boundary shape as Φ_R , the region-based external energy term is defined using voxels within a narrow band

around the model surface as: $E_R = \int_A \Phi_A(\mathbf{v})\Phi_R(\mathbf{v})dA$. The multiplicative term provides two-way balloon forces that deform the model toward the predicted ROI boundary. This allows flexible model initializations either overlapping the object or inside the object. Using these external energy terms as image forces, the FEM model is driven to find object boundaries.

Hierarchical Shape Priors and Model Evolution: The above model may still not be able to avoid the local minimum or preserve a specific shape, especially when the texture of the image is complex. Thus a shape prior energy is added to our method to constrain and refine the shape and position of all n structures during deformation: $E_{prior} = \sum_{i=1}^n E_{local}^i + k_g \cdot E_{global}$, where E_{local}^i is the shape constraint applied on the i th structure, and E_{global} is a constraint on the relative positions among different structures and distances with respect to the mass centroid.

In terms of the implementation, the energy optimization is achieved by using an expectation-maximization (EM) type of framework. Given intermediate segmentation results of all structures, in the ‘‘E’’ step, the locations of structures are adjusted as per the global statistics to prevent overlap or intersection. In the ‘‘M’’ step, each individual shape is constrained by aligning to the mean shape, and then mapping into PCA space to update the pose and shape parameters. Thus it guarantees that the shape deforms only into patterns consistent with the training data, which can refine the shape and prevent over-segmentation. This whole hierarchical framework is employed adaptively by defining $k_g = 1 - e^{-|\nabla A|}$, where $|\nabla A|$ is the magnitude of deformation change in surface shape. In the beginning, the shape deforms a lot so k_g is relatively large. It means that we trust more on the high level information, and put larger weights on the global statistics. After several iterations, the local statistics should be more important since the brain structure is nearby the boundary. At this time the shape deforms less so k_g is smaller. The whole segmentation framework, including training and testing, is summarized as follows:

1. Manually segment a small number of training data. Use PCA to capture shape statistics and location statistics of relative positions.
2. Given a testing data, it is aligned to the reference image. Then multiple models are placed using mean relative positions with respect to the mean mass centroid. Initialize these models, i.e., stiffness matrix and step size for FEM and the gradient magnitude or edge map.
3. Compute Φ_A based on the current model; predict object ROI R by applying the Bayesian Decision rule to binarizing the estimated object probability map, and compute Φ_R . Calculate the external force vector.
4. Deform the model using FEM and external forces derived from gradient and region terms. Smoothness constraint is implicitly incorporated by FEM.
5. Update the value of $k_g = 1 - e^{-|\nabla A|}$. Adjust the structure positions to prevent intersections using global prior.
6. Refine each intermediate result by local prior constraint (transform it to the mean shape, then update the pose and shape parameters).
7. Repeat steps 3-6 until convergence.

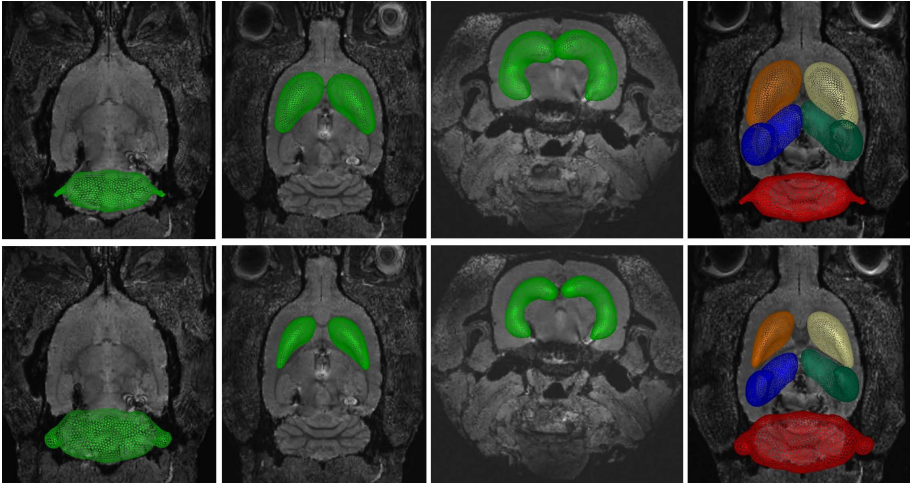


Fig. 3. Comparison of segmentation results starting from the same initialization. The first row: traditional shape prior constraint (smoothness). The second row: using hierarchical shape prior. From left to right: the cerebellum, the left and right striatum, the left and right hippocampus, and all five structures.

Alternating the deformation module and the shape constraint module is more robust to noise and can handle more complex textures than purely using a deformation module. Due to the constraint from the shape prior, our model converges fast and robustly towards the true boundary given the mean shape of the multiple objects as initialization. One more benefit of our multiple structure model is that the size of the training data can be small, because the shape prior is built hierarchically. The prior for each structure is obtained individually, which can better discover the shape statistics than using one prior for all structures.

3 Experiments

Experimental Settings: The heads of adult male Sprague-Dawley rats were scanned on a 21.1T Bruker Biospin Avance scanner. The protocol consisted of a 3D T2-weighted scan with echo-time (TE) $7.5ms$, repetition time (TR) $150ms$, 27.7 kHz bandwidth, field of view (FOV) of $3.4 \times 3.2 \times 3.0mm$, and voxel size $0.08mm$, isotropic. 3D annotation is manually performed on 58 volume data by an expert neurologist. 8 are used as training data since we only need a small number of samples for hierarchical statistics. The rest 50 are used as testing. The proposed method was implemented in C++ and Python 2.6 and tested on a 2.40 GHz Intel Core2 Quad computer with 8G RAM.

Visual comparisons: Fig. 3 shows the visual comparisons of segmentation results starting from the same initialization. In the first row, we segment each

structure individually. The shape prior is the smoothness constraint. In the second row, the hierarchical shape prior energy is incorporated to segment all structures simultaneously. When segmenting the cerebellum, the shape prior helps to preserve the shape of paraflocculi and to avoid the local minima caused by complex textures inside. Without using our priors, the model is attracted by large gradient values inside and the paraflocculi are smoothed out. In the striatum and hippocampus cases, there is no obvious boundary between these structures and the brain. Thus the model can easily over-segment the ROI. Our priors alleviate this problem. Another benefit of the hierarchical prior is that the spatial constraint is incorporated. The distance between two structures has to follow the location statistics. Thus structures cannot intersect with each other.

Quantitative comparisons: Tab. 1 compares different priors. Note that all deformation modules are the same and based on both gradient and region information. We compared the methods of combining the robust deformation model with a traditional smoothness shape prior [10], an independent shape prior [12] for each structure and hierarchical shape priors (i.e., the proposed method). The same parameters are used in all deformation modules. We reported the mean value and standard deviation of voxel distances between segmented surfaces and ground truth surfaces. Generally, the shape prior constraint improves the segmentation accuracy. In the cerebellum case, the spatial constraint in hierarchical priors only slightly improves the result, while it is much more important in the striatum and hippocampus cases. The reason is that the striatum and hippocampus are adjacent to each other. Spatial constraints can prevent the intersection of different structures, which implicitly alleviates the over-segmentation problem caused by the low contrast and ambiguous boundaries. The volume magnitude information is very important in rodent brain analysis. Thus we also reported the relative error of volume magnitude compared to the ground truth. Using the hierarchical shape prior achieves the best result, especially for the striatum and the hippocampus. This spatial constraint not only improves the segmentation accuracy for multiple structures, but also decreases the number of iterations and computation time to converge.

Table 1. Quantitative comparisons of different prior schemes. We reported the mean and standard deviation of voxel distances between segmented surfaces and ground truth surfaces, and relative errors of volume magnitude in proportions

Structures	Prior types	Voxel distance	Volume error
Cerebellum	Smoothness	4.35 ± 2.17	0.22 ± 0.12
	Independent	1.74 ± 1.18	0.05 ± 0.02
	Hierarchical	1.70 ± 1.13	0.04 ± 0.02
Striatum	Smoothness	3.79 ± 2.05	0.51 ± 0.19
	Independent	2.93 ± 1.81	0.19 ± 0.06
	Hierarchical	1.37 ± 1.09	0.07 ± 0.03
Hippocampus	Smoothness	3.82 ± 2.14	0.53 ± 0.18
	Independent	2.69 ± 1.83	0.17 ± 0.05
	Hierarchical	1.22 ± 1.05	0.06 ± 0.02

4 Conclusions

In this paper, we proposed a deformable segmentation method to segment multiple rodent brain structures simultaneously. The model is driven by external forces derived from both image gradient and appearance information. To improve the robustness, a hierarchical shape prior module is also employed for model initialization and shape refinement. We use this method to segment complex structures in rodent brains, which shows that our method works effectively.

References

1. Chen, T., Rangarajan, A., Eisenschenk, S., Vemuri, B.: Construction of neuroanatomical shape complex atlas from 3D brain MRI. In: Jiang, T., Navab, N., Pluim, J., Viergever, M. (eds.) MICCAI 2010, Part III. LNCS, vol. 6363, pp. 65–72. Springer, Heidelberg (2010)
2. Cootes, T., Taylor, C., Cooper, D., Graham, J.: Active shape model - their training and application. *Computer Vision and Image Understanding* 61, 38–59 (1995)
3. Davatzikos, C., Tao, X., Shen, D.: Hierarchical active shape models, using the wavelet transform. *IEEE Transactions on Medical Imaging* 22(3), 414–423 (2003)
4. Frangi, A., Rueckert, D., Schnabel, J., Niessen, W.: Automatic construction of multiple-object three-dimensional statistical shape models: Application to cardiac modeling. *IEEE Transactions on Medical Imaging* 21(9), 1151–1166 (2002)
5. Huang, X., Metaxas, D.: Metamorphs: Deformable shape and appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30(8), 1444–1459 (2008)
6. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. *International Journal of Computer Vision* 1, 321–331 (1987)
7. Malladi, R., Sethian, J., Vemuri, B.: Shape modeling with front propagation: A level set approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 17, 158–175 (1995)
8. Paragios, N., Deriche, R.: Geodesic active regions and level set methods for supervised texture segmentation. *International Journal of Computer Vision* 46(3), 223–247 (2002)
9. Shen, D., Davatzikos, C.: An adaptive-focus deformable model using statistical and geometric information. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22(8), 906–913 (2000)
10. Shen, T., Li, H., Qian, Z., Huang, X.: Active volume models for 3D medical image segmentation. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 707–714 (2009)
11. Staib, L., Duncan, J.: Boundary finding with parametrically deformable models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14(11), 1061–1075 (1992)
12. Zhang, S., Huang, J., Uzunbas, M., Shen, T., Delis, F., Huang, X., Volkow, N., Thanos, P., Metaxas, D.: 3D segmentation of rodent brain structures using active volume model with shape priors. In: *ISBI* (2011)
13. Zhu, S., Yuille, A.: Region Competition: Unifying snakes, region growing, and Bayes/MDL for multi-band image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 18(9), 884–900 (1996)

Digital Imaging for the Education of Proper Surgical Hand Disinfection^{*}

Tamás Haidegger¹, Melinda Nagy¹, Ákos Lehotsky¹, and László Szilágyi²

¹ Budapest University of Technology and Economics, Dept. of Control Engineering and Information Technology, Budapest, Hungary

haidegger@iit.bme.hu, meliole28@gmail.com, akos.lehotsky@gmail.com

² Sapiientia - Hungarian Science University of Transylvania,
Faculty of Technical and Human Science, Tirgu-Mures, Romania
lalo@ms.sapiientia.ro

Abstract. Nosocomial infections are the undesirable result of a treatment in a hospital, or a health care service unit, not related to the patient's original condition. Despite the evolution of medicine, fundamental problems with hand hygiene remain existent, leading to the spread of nosocomial infections. Our group has been working on a generic solution to provide a method and apparatus to teach and verify proper hand disinfection. The general idea is to mark the skin surfaces that were sufficiently treated with alcoholic hand rub. Digital image processing is employed to determine the location of these areas and overlay it on the segmented hand, visualizing the results in an intuitive form. A non-disruptive ultraviolet marker is mixed to a commercially available hand rub, therefore leaving the original hand washing workflow intact. Digital images are taken in an enclosed device we developed for this purpose. First, robust hand contour segmentation is performed, then a histogram-based formulation of the fuzzy *c*-means algorithm is employed for the classification of clean versus dirty regions, minimizing the processing time of the images. The method and device have been tested in 3 hospitals in Hungary, Romania and Singapore, on surgeons, residents, medical students and nurses. A health care professional verified the results of the segmentation, since no gold standard is available for the recorded human cases. We were able to identify the hand boundaries correctly in 99.2% of the cases. The device can give objective feedback to medical students and staff to develop and maintain proper hand disinfection practice.

Keywords: surgical hand disinfection, hospital acquired infection, image segmentation, semi-supervised fuzzy *c*-means clustering.

^{*} This work was supported in part by Contact Singapore, the Austrian Center for Medical Innovation and Technology and the Hungarian national grants NKTH OTKA CK80316 and TÁMOP-4.2.1/B-09/1/KMR-2010-0002. The work of L. Szilágyi was supported by CNCSIS-UEFISCSU, project no. PD_667, contract no. 28/05.08.2010.

1 Introduction

Ignaz Semmelweis realized first the importance of hand hygiene in the medical environment in 1847, and imposed strict regulations in his hospital, immediately resulting in better patient survival rates. While medical technology has tremendously improved along the years, it has failed to eliminate the persistent problem of incomplete hand hygiene, leading to the spread of nosocomial infections. Infections are considered nosocomial—also referred to as Hospital-Acquired Infections (HAI)—if they appear 48 hours or more after hospital admission or within 30 days after discharge. HAI occurrence has been reported in 7.1% on average in the Western world [1]. Sadly, HAI is responsible for 15,000 deaths in Europe and over 90,000 world wide every year [2]. In addition, it causes \$2B extra cost annually in the UK [3], \$4.5–5.7B direct [4] and \$17–40B additional costs in the USA, as it may triple average hospitalization time [3,5].

The pathogens are most often transferred via the hands of the medical staff, therefore compliance with the operative hand hygiene protocols and standards is absolutely necessary [6,7,8]. Fig. 1(a) shows the most frequently missed parts of the hands based on a multi-center study conducted by the US Center of Disease Control and Prevention (CDC) [9]. Reports suggest that accurate hand disinfection could reduce HAI by up to 30%, saving hundreds of lives each month and reducing hospitalization costs significantly [10,11].

In this paper, a novel approach to hand disinfection education and quality control is discussed. The employed workflow is very intuitive: disinfection is performed with a liquid, alcohol-based hand rub mixed with UV reflective powder, which appears brighter under adequate UV-C lighting (on surfaces having sufficiently treated with the soap). Pictures of the hands are taken in a UV lighted case (users should spread their fingers to provide good visual access), after which a regular notebook performs the evaluation. It provides graphical results (highlighting clean versus dirty areas), and indicates an overall quantitative score. We can obtain objective, measurable and scalable data on hand disinfection quality, that is most useful to train medical students to acquire good hand washing practice.

2 Methods

2.1 Image Recording

A compact, mobile device was built for comparable recording and objective assessment of hand disinfection quality, shown in Fig. 1(b). This consists of removable plastic plates, making it light and transportable. Dimensions are $35 \times 30 \times 30$ cm (w/l/h) providing dual hand imaging option. There are 21 lines with 10 UV LEDs in each row, fixed on the top plate. LEDs provide equal illumination on the hand surface. The current is minimized to 30 mA, therefore any USB port is sufficient for power supply, not constraining the mobility of the device. In the center of the cover, a digital camera (Canon PowerShot SX100) is placed, controlled through the connected laptop. The whole structure is covered

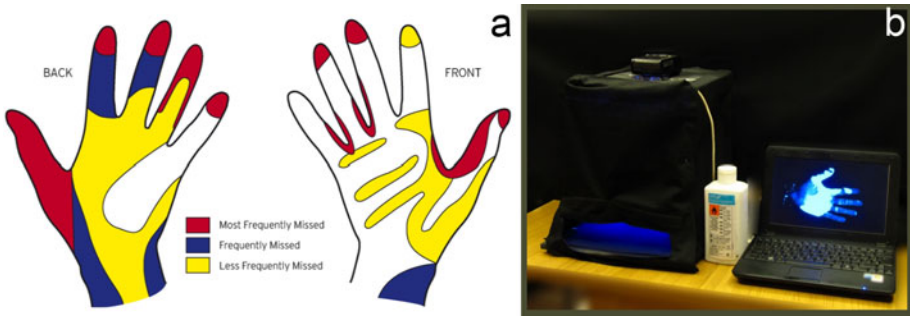


Fig. 1. (a) Results of a study by CDC, showing the most frequently missed parts of the hands after surgical hand disinfection, based on a US survey [9]; (b) Hardware setup for objective hand disinfection control

with removable drapes, reducing the disturbance of external light. The compact device was designed to be wall-mountable for upcoming prolonged (repetitive) clinical trials. In order to keep the interior of the box clean, we use disposable matte black paper covers.

2.2 Initial Considerations for Image Processing

The aim of the system is to provide an objective and repeatable measurement of hand washing through image segmentation and classification by determining the ratio of the clean (bright) surface compared to the whole skin surface. To be able to compute this, the Region of Interest (ROI) is first selected, separating the hand from the background. Next, the pixels within the ROI are fed to a semi-supervised classification process, to distinguish clean areas from dirty ones. Finally, the ratio of the clean areas over the whole hand is calculated.

For medical applicability, it is important to find a good compromise between processing speed and image resolution. Accordingly, captured images are resized to 400×300 pixels. Although the camera captures three intensity channels (red, green and blue), we restricted the procedure to use the green channel, since the red channel's intensity is very low in the reduced-light environment, therefore it rarely contains less information. On the other hand, during the initial trials, it was noticed that the blue channel is often saturated.

2.3 ROI Selection

The first step of the procedure mostly consists of image processing operations that can be easily fed to parallel execution:

1. A median filter of 5×5 mask is applied to get rid of isolated pixels containing high-frequency imaging artifacts.

2. The complex morphological operation is applied to the filtered image:

$$I_g = [(I_f \oplus b) - (I_f \ominus b)] \times f(I_f \ominus b) , \quad (1)$$

where b is a 3×3 square shaped structuring element, I_f stands for the filtered image, I_g represents the computed gradient image, and $f(\cdot)$ is an amplification function. The role of $f(\cdot)$ is to amplify the edges that are situated in the proximity of dark pixels in the filtered image, and suppress other edges, which are supposed to be inside the area of the hand. $f(\cdot)$ is an exponentially decreasing function that crosses the unit value in a context dependent threshold intensity level.

3. In gradient image I_g , the background typically contains visible texture, while the inner area of the hand is uniformly dark. A region growing method is applied to identify the pixels belonging to the hand. In order to select seed points, a uniform 11×11 averaging mask is applied to the image I_g . Minimum intensity pixels of the smoothed image can be used as seed points.
4. Finally, the pixels belonging to the amplified contour of the hand are separated, in accordance with their original intensities.

The intermediary and final results of ROI selection are shown in Fig. 2(a)–(f). After the ROI is established, the relative histogram of the inner pixels is computed, using the intensities of the filtered image, I_f . This histogram (see p_l values in Section 2.4) will be the input data of the semi-supervised classification.

2.4 Semi-supervised Classification

To minimize the processing time of the images, a quick, histogram-based formulation [12] of the fuzzy c -means algorithm (FCM) [13] is employed for the separation of pixels belonging to clean and dirty regions of the hand. The main reason for this approach is its ability to classify intensity levels instead of pixels, resulting in high execution speed. Supervision is introduced into the objective function of FCM by the second additive term, and its strength is controlled by the supervision rate σ that acts like a trade-off parameter between the unsupervised and supervised terms. The proposed objective function is:

$$J_{\text{ssFCM}} = (1 - \sigma) \sum_{i=1}^c \sum_{l=1}^q p_l u_{il}^m (l - v_i)^2 + \sigma \sum_{i=1}^c \sum_{l=1}^q \tilde{p}_{il} (l - v_i)^2 , \quad (2)$$

where c represents the number of clusters, q is the number of intensity levels in the input image, v_i is the prototype intensity of cluster with index i , u_{il} is the fuzzy membership function representing the probability of gray shade l to be assigned to cluster represented by v_i and $m > 1$ is the fuzzy exponent as defined in [13]. Further on, p_l is the percentage of pixels situated in the ROI having their intensity equal to l and similarly, \tilde{p}_{il} , $l = 1 \dots q$ describe the intensity distribution of the labeled data belonging to cluster index i . These previously labeled pixels were extracted from several expert-evaluated reference images, and they are assigned to clusters using hard (bivalent) partition.

The objective function J_{ssFCM} is optimized under the probabilistic constraint $\sum_{i=1}^c u_{il} = 1, \forall l = 1 \dots q$. The optimum is reached by alternately optimizing along u_{il} with v_i fixed and along v_i with u_{il} fixed, until the cluster prototypes converge. The update formulas of fuzzy memberships and cluster prototypes—obtained using zero gradient conditions and Lagrange multipliers—are:

$$u_{il} = \frac{(l - v_i)^{-2/(m-1)}}{\sum_{j=1}^c (l - v_j)^{-2/(m-1)}}, \quad \forall i = 1 \dots c, \forall l = 1 \dots q, \quad (3)$$

and

$$v_i = \frac{\sum_{l=1}^q [(1 - \sigma)p_l u_{il}^m + \sigma \tilde{p}_{il}] \times l}{\sum_{l=1}^q (1 - \sigma)p_l u_{il}^m + \sigma \tilde{p}_{il}}, \quad \forall i = 1 \dots c. \quad (4)$$

Clustering is executed with the class count initially set to $c = 3$. Besides the clean and dirty areas, an intermediate (so-called partially washed) area is also identified. However, depending on the final intensity values of the cluster prototypes, this middle cluster may be merged into the dirty or clean class, due to excessive proximity. In the next phase of the development, we intend to involve patient-specific data (skin color, vessels, birthmarks, etc.) to facilitate the final decision on the intermediate class.

2.5 Interpretation of Classification

At this stage, pixels are clustered regardless of their position within the ROI. Next, we apply certain context-dependent criteria to improve the prediction accuracy. A regularization filter is used to remove isolated pixels of any class, surrounded by another class. This also helps to remove small spots that originate from uneven skin color. Furthermore, a weighting factor is assigned to each pixel within the ROI (with lower weight to pixels situated along the edges), to minimize the influence of shadows on the final estimation. The weighted summation of correctly disinfected areas of the hand is achieved through:

1. A distance transform is employed to separate the pixels within the ROI into edge pixels and inner pixels. Edge pixels are situated within a close neighborhood of the non-hand regions. We denote by n the number of edge pixels, while the number of inner pixels to be referred to as N .
2. We differentiate and count clean and dirty pixels within the set of edge pixels (n_c and n_d will denote the count of clean and dirty edge pixels, respectively) and inner pixels (N_c and N_d , analogously).
3. The non-weighted estimation of the clean area is given as:

$$\pi = 100 \times \frac{N_c + n_c}{N_c + n_c + N_d + n_d}. \quad (5)$$

4. The weighted (edge-compensated) percentage of the disinfected area is:

$$\pi_w = 100 \times \frac{N_c + \alpha n_c}{(N_c + N_d) + \alpha (n_c + n_d)}. \quad (6)$$

5. In order to suppress shading artifacts, an $\alpha = 0.2$ weighting factor is recommended. This value was established empirically.

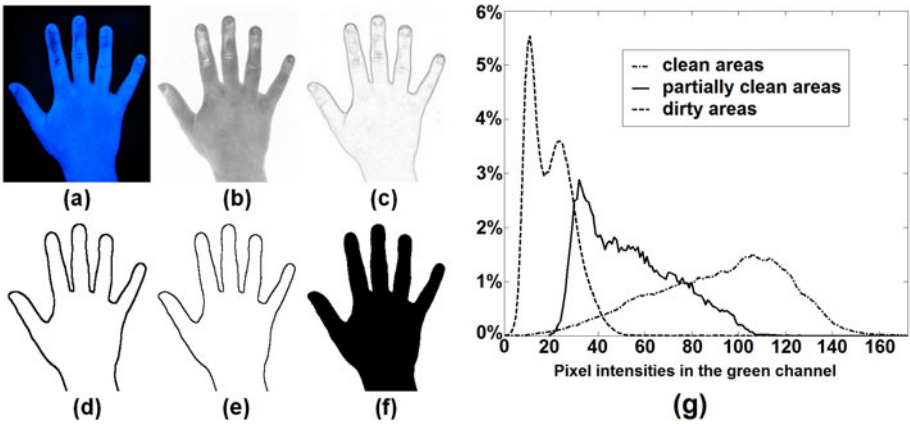


Fig. 2. (a)-(f) Intermediary stages of hand segmentation: (a) original image, (b) green channel image, (c) gradient of the green channel image, (d) initial detected contour, (e) adjusted contour, (f) extracted ROI (images (b)–(f) are shown in inverse colors); (g) Relative histogram of labeled data involved in the semi-supervised clustering algorithm

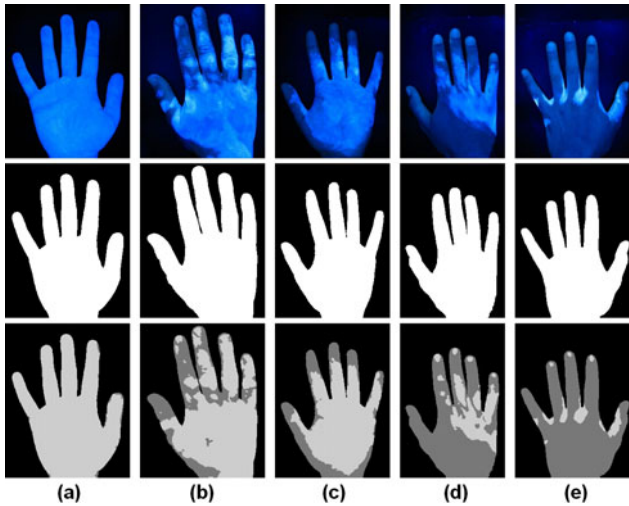


Fig. 3. Qualitative evaluation results for visual inspection: (a)–(e) hands with total or partial disinfection. In each column: upper row shows original color image, middle row indicates the extracted ROI, bottom row exhibits the segmented image.

3 Results and Discussion

To test the efficiency of the system, a trilateral study was conducted. Medical workers including surgeons, nurses, students and supporting staff from different

Table 1. Quantitative results regarding the clean areas of the hands from Fig. 3

Hand	(a)	(b)	(c)	(d)	(e)
Before compensation (π)	98.72%	71.37%	72.12%	24.62%	8.73%
After compensation (π_w)	99.47%	73.90%	75.90%	25.73%	8.41%

hospitals volunteered to take part in the trials. Institutes were selected from Budapest (Hungary), Tîrgu-Mureş (Romania) and Singapore.

In an initial phase, approximately 200 images were recorded from dozens of different subjects, intentionally having varying disinfected area ranging from 0% to 100%, forming a generalized teaching set.

Detected boundaries underwent a double investigation: first a visual inspection was applied, followed by an automatic evaluation of the pixel intensities in the immediate neighborhood of the detected boundary. This latter criterion approved the boundary when the great majority (e.g. 90%) of inner neighbor pixels had higher intensity (in the green channel of the original image) than the great majority of outer neighbor pixels. In 99.2% of the tested cases, the boundary of the hand was correctly identified. The intermediary and final result of a successful case are presented in Fig. 2(a)–(f). The labeled data for supervised clustering was manually selected by experts from a learning data set that consisted of 64 different images. The relative histogram of labeled pixels belonging to the three classes is shown in Fig. 2(g). The three classes have significant overlapping parts.

The algorithm was tested on several images with different supervision rates (σ) ranging between 0.005 and 0.03, and the optimal value of σ was derives to be 0.015. Fig. 3 exhibits the result of five successful segmentations; identified clean areas are drawn in lighter gray, while dirty parts appear in darker gray. Table 1 gives the results of the quantitative evaluation of the same five images. As a guideline for future applications, we consider a hand appropriately disinfected, when it resembles the case shown in Fig. 3(a), having the final estimated percentage greater than 98%.

During the test of the system, one of the challenges experienced was to account for the skin differences. Using calibrated image recording conditions (like foxed exposure time, ISO value of the camera) and including labeled data into the clustering algorithm resolved this problem with a remarkable accuracy.

The proposed image processing algorithm was implemented in C++ programming language, using Intel’s OpenCV computer vision libraries. Image processing tasks rarely handle individual pixels, most operations apply to the whole image (possibly restricted by a ROI), thus facilitating parallel execution. This enables us to provide the final evaluation results of an image in less than 1.5 seconds.

4 Conclusions

Since insufficient sterility continues to cause serious problems in health care around the globe, we exerted significant efforts to support the prevention of Hospital-Acquired Infections. We have built a mobile device that can perform the

objective evaluation of the hand's disinfection ratio. The equipment is intended to help surgeons and hospital staff to reduce HAI rates, through teaching better hand hygiene practice. We use a UV reflective powder mixed to the regular alcoholic hand rub, and then clean areas of the hand are identified from digital images taken under UV lighting. The main contribution was the creation of an automated device and adjoint procedure that efficiently and accurately segments the digital image of the hand, and then classifies the treated areas using a semi-supervised fuzzy *c*-means clustering algorithm. The performance of the system was evaluated on a large set of images, leading to an overall accuracy of 98.5%, as verified by a human professional. It has been shown in trilateral clinical trials that the system is qualified for prevailing the omitted parts of the hand after hand washing. Meanwhile, it does not significantly disrupt the regular medical workflow, and is also capable of tracking disobedience to hand hygiene protocol, such as wearing jewelry or artificial nails.

References

1. European Center for Disease Prevention and Control: Annual Epidemiological Report 2008: Focus on Healthcare-associated infections, special chapter, pp. 16–38 (2008)
2. Pollack, A.: Rising threat of infections unfazed by antibiotics, *New York Times* (February 26, 2010)
3. Plowman, R., Graves, N., Griffin, M., Roberts, J.A., Swan, A.V., Cookson, B., Taylor, L.: The socio-economic burden of hospital acquired infection. Public Health Laboratory Service, London (1999)
4. Korniewicz, D.M.: Hand hygiene compliance and healthcare-associated infections—Who is responsible? *Touch Briefings—Infection Control*, 37–39 (2007)
5. Xhale Technologies Inc.: HyGreen, <http://www.xhale.com/hygreen>
6. Girou, E., Loyeau, S., Legrand, P., Oppein, F., Brun-Buisson, C.: Efficacy of handrubbing with alcohol based solution versus standard handwashing with antiseptic soap: randomised clinical trial. *Brit. Med. J.* 325, 362–365 (2002)
7. Goroncy-Bermes, P.: Hand disinfection according to the European Standard EN 1500 (hygienic handrub): a study with Gram-negative and Gram-positive test organisms. *Int. J. Hyg. Environ. Health* 204(2-3), 123–126 (2001)
8. Boyce, J.M., Pittet, D.: Guideline for hand hygiene in healthcare settings: recommendations of the HICP Advisory Committee and HICPAC/SHEA/APIC/IDSA Hand Hygiene Task Force. *Morbidity&Mortality* 51(RR16), 1–44 (2002)
9. CDC Joint Commission: Measuring hand hygiene adherence: overcoming the challenges. Oakbrook Terrace, IL (2009)
10. Pittet, D., Mourouga, P., Perneger, T.V.: Compliance with handwashing in a teaching hospital. *Ann. Inter. Med.* 130(2), 126–130 (2002)
11. WHO Collaborating Centre for Patient Safety Solutions: Improved hand hygiene to prevent health care-associated infections. *Patient Safety Solutions* 1(9), 1–4 (2007)
12. Szilágyi, L., Benyó, Z., Szilágyi, S.M., Adam, H.S.: MR brain image segmentation using an enhanced fuzzy *c*-means algorithm. In: 25th Annual Int'l. Conference of IEEE Engineering in Medicine and Biology Society, pp. 724–726 (2003)
13. Bezdek, J.C.: Pattern recognition with fuzzy objective function algorithms. Plenum, NY (1981)

Reinforcement Learning for Context Aware Segmentation

Lichao Wang, Robert Merrifield, and Guang-Zhong Yang

The Hamlyn Centre for Robotic Surgery, Imperial College London, UK
{lichao.wang, rdm99, g.z.yang}@imperial.ac.uk

Abstract. The ability to learn from user behavior during image segmentation to replicate the innate human ability to adapt shape delineation to contextually specific local information is an important area of study in image understanding. Current approaches to image segmentation usually incorporate specific designs, either relying on generic image features or specific prior knowledge, which usually prevent their application in different contextual settings. In this paper, a general segmentation framework based on reinforcement learning is proposed. It demonstrates how user-specific behavior can be assimilated *in-situ* for effective model adaptation and learning. It incorporates a two-layer reinforcement learning algorithm that constructs the model from accumulated experience during user interaction. As the algorithm learns ‘pervasively’ whilst the user performs manual segmentation, no additional steps are required for the training process, allowing the method to adapt and improve its accuracy as experience is acquired. Detailed validation of the method on *in-vivo* magnetic resonance (MR) data demonstrates the practical value of the technique in significantly reducing the level of user interaction required, whilst maintaining the overall segmentation accuracy.

1 Introduction

Automated image segmentation for medical imaging is inherently challenging due to the complexity of anatomical structures and inevitable artefacts introduced during data acquisition. Current approaches can be divided into two main categories: methods relying on generic image features and those combining prior knowledge, whether being geometrical or statistical, with specific image content being examined. Active contours and level-set methods [1, 2] are typical approaches focusing on the current image content, whereas Statistical Shape Models (SSM) [3] instantiate the overall behavior of a group of shapes on a specific data set. Techniques such as SSM are heavily dependent on a comprehensive training set with residual issues of over-fitting and poor generalisation to pathological data in clinical application [4-6]. Thus far, limited attention is paid to the learning of user behavior during manual segmentation. It has long been recognised that experienced observers have innate abilities in adapting shape delineation to local information that is contextually specific. Fig. 1 illustrates three typical examples of image segmentation showing (a) the endocardial border of the left ventricle (LV), where users do not always rely on gradient features for myocardial mass measurement when papillary muscles are excluded; (b) the

avoidance of the partial volume effect of MR; and (c) bridging across both subtle and strong anatomical boundaries during the segmentation of the inner and outer contours of the levator ani, the tripartite muscle sheet of the pelvic floor.

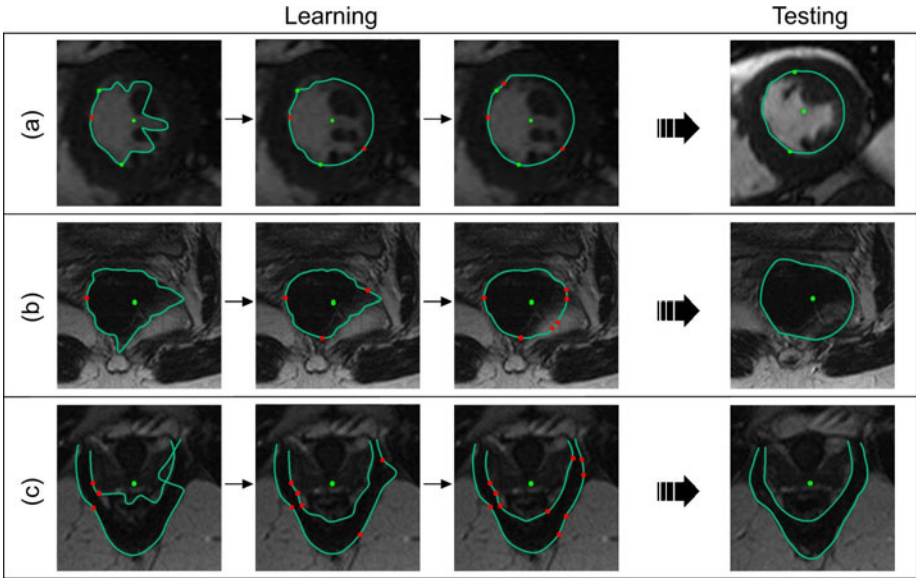


Fig. 1. Segmentation examples illustrating how user delineation or modification of the anatomical contours (shown in red) can be learnt through reinforcement learning and then used to perform context-specific segmentation: (a) the endocardial border of the left ventricle when the papillary muscles are excluded; (b) the boundary of the rectum showing how to adapt to the partial volume effect in MR; (c) the inner and outer contours of the pelvic floor when both subtle and strong appearance features are present. It is worth noting that the algorithm learns to extrapolate the shape when there are no clear image appearance features to rely upon. Under the proposed reinforcement learning framework, the predicted boundary points get either reward or punishment through user acceptance (implicit) or correction (explicit), thus making the segmentation model context specific and generalizable to unseen geometrical variations.

The purpose of this paper is to propose a reinforcement learning framework for context aware segmentation and to demonstrate how user-specific behavior can be assimilated *in-situ* for effective model adaptation and learning. Central to the proposed technique is a two-layer reinforcement learning algorithm that builds the model from accumulated experience during user interaction. One advantage of the framework is that the learning process circumvents the need for extensive off-line training, and thus is readily generalizable to unseen pathological patient cases in a context or subject specific manner. As the algorithm learns ‘*pervasively*’ when the user performs manual segmentation, no additional steps are required for the training process. Furthermore, the method is able to adapt and improve its accuracy as experience is acquired, a process that is akin to the training of human observers. The example results in Fig. 1 demonstrate the potential of the technique. In this paper, we use left ventricular segmentation from MR images to illustrate the overall structure of the rein-

forcement learning framework and assess the extent to which the level of user interaction can be reduced, thus leading to an efficient semi-automatic segmentation approach.

2 Methods

2.1 Two-Layer Reinforcement Learning

Reinforcement learning [7] is a process where the learning agent interacts with the environment and tries to maximize its expected long-term reward. It is particularly suited to problems where explicit knowledge for solutions is difficult to obtain. By interacting with the user, the learning agent is not informed explicitly how to combine different strategies to attain the correct contour point. Instead, with the example given by the user, the agent only knows the level of its own performance. In general, there are three key elements of reinforcement learning, namely: state, action and reward. A state is a situation of the environment. An action is some behavior of the learning agent that interacts with the environment. A reward is the feedback that the agent receives from the environment during interaction, and in the case of image segmentation, the quality of localizing anatomical structures. The expected long-term reward of a given state or state-action pair is referred to as a *value*.

Q-learning [8] is a widely used reinforcement learning algorithm, where a table is maintained for recording values of all the possible states or state-action pairs. For image segmentation, however, due to the diversity of both geometrical and appearance information present within the image, inevitably there tend to be many states and actions for the learning agent; therefore the standard Q-learning approach in this case can be problematic. One way to solve this problem is to use value function approximation. To perform the approximation, parameters with much lower dimensionality than the actual state-action space are used to describe the value function. With these parameters, the agent does not attempt to find the best action for a given state directly as in Q-learning. Instead, it estimates the parameters that maximize the value function, which correspond to the best action. The estimation can be performed by minimizing the mean square error over the observed examples. For the minimization process, a policy gradient approach can be used. This adjusts the parameter vector during trials along the direction where the estimation error is minimized for the given example:

$$\bar{\theta}_{t+1} = \bar{\theta}_t + \gamma [V^\pi(a_t) - V_t(a_t)] \nabla_{\theta} V_t(a_t) \quad (1)$$

where $\bar{\theta}$ is the parameter vector, t is the step index, $V^\pi(a_t)$ is the true value of the action a_t , $V_t(a_t)$ is the estimate value at step t , and γ is a constant coefficient.

For a typical segmentation task, the three elements of the reinforcement learning framework can be defined as follows: a state is the appearance (e.g. the intensities along a given search direction); an action for a state is to localize a point on the anatomical border; and the reward is defined as an evaluation function. The parameters for segmentation are the weightings of different candidate strategies used to localize contour points.

For the purpose of left-ventricular segmentation, candidate strategies include an Akima spline [9] as the basic geometrical representation of the anatomy due to its resilience to outliers. For simplicity, we have adopted a radial search strategy commonly used for LV segmentation. By combining both appearance and geometry information, the cost function to localize a contour point on a given search direction can be defined as:

$$f(x, \alpha, \bar{w} | I) = \alpha \left(\sum_{i=1}^n w_i f_{A_i}(x | I) \right) + (1 - \alpha) f_G(x) \quad (2)$$

where x is the pixel position on the radial, I is the intensity vector, $f_{A_i}(x)$ is the i th strategy within the appearance strategy category, w_i is weighting for $f_{A_i}(x)$, or referred to as the i th in-category weighting, α is the weighting for the appearance strategy category, or referred to as cross-category weighting, $f_G(x)$ is the cost function for the geometry strategy, the weighting vector $\bar{w} = (w_1, w_2, \dots, w_n)^T$ and $\sum_{i=1}^n w_i = 1$.

The candidate contour point is defined by minimizing Eq. (2). This is denoted by x_{\min} and is a function of weightings. Given an arbitrary set of weightings, x_{\min} may not be optimum in terms of segmentation accuracy. The distance to the user example's position on the same radial can be used for evaluating x_{\min} :

$$g(\alpha, \bar{w} | I, f_{A_1}, \dots, f_{A_n}, f_G) = - \left| x_{\min}(\alpha, \bar{w} | I, f_{A_1}, \dots, f_{A_n}, f_G) - x_T \right| \quad (3)$$

where x_T is the example position. Therefore the learning task is, for every example, to optimize Eq. (3) with respect to the weightings. After learning, the optimal weightings are obtained for explicitly learnt samples, typically via user interaction.

In this paper, learning is separated into two layers, *i.e.*, in-category weighting (*i.e.*, \bar{w}) learning for appearance strategies, and cross-category weighting (*i.e.*, α) learning for appearance and geometry categories. This corresponds to a two-stage process, *i.e.*, firstly to use appearance to localize the clear contour areas, and if it fails, then to combine geometry to interpolate the shape. Hence the evaluation function represented by Eq. (3) can be sub-divided into the following equations:

$$g_{in}(\bar{w} | I, f_{A_1}, \dots, f_{A_n}) = - \left| x'_{\min}(\bar{w} | I, f_{A_1}, \dots, f_{A_n}) - x_T \right| \quad (4)$$

$$g_{cross}(\alpha | \bar{w}, I, f_{A_1}, \dots, f_{A_n}, f_G) = - \left| x_{\min}(\alpha | \bar{w}, I, f_{A_1}, \dots, f_{A_n}, f_G) - x_T \right| \quad (5)$$

Eq. (4) is the evaluation of the localized point using appearance strategies only (x'_{\min}), whereas Eq. (5) is the evaluation of using both appearance and geometry strategies.

In the context of in-category weighting learning, the parameters for function approximation are the weightings for different appearance-based strategies. Meanwhile, because every learning episode for point localization on a radial has only one step, the immediate reward is the same as the expected long-term reward. Thus by replacing

θ with \vec{w} , $V^\pi(a_t)$ with g_m from Eq. (4) and $V_t(a_t)$ with 0 (since the learning agent expects its current action has zero error), Eq. (1) becomes

$$\vec{w}_{t+1} = \vec{w}_t + \gamma \left| (g_m(\vec{w}))_t \right| \nabla_{\vec{w}} (g_m)_t(\vec{w}_t) \quad (6)$$

In practice, the parameter update is realized as follows: when an example contour point is given by the observer, the algorithm regards it as the ground truth position on the contour; it then calculates the local gradient of the in-category weightings of appearance strategies and updates the weightings using Eq. (6), trying to minimize the difference between its localization and the user example.

For cross-category weighting learning, the weighting update equation is similar to (6) using policy gradient:

$$\alpha_{t+1} = \alpha_t + \gamma \left| (g_{cross}(\alpha))_t \right| \nabla_{\alpha} (g_{cross})_t(\alpha_t) \quad (7)$$

However, the learning flow for the cross-category weighting is significantly different due to two reasons. Firstly, unlike f_{A_i} in Eq. (4) which are independent from each other, f_G in Eq. (5) depends on the appearance strategies, i.e., f_{A_i} and \vec{w} . This dependency is due to the design that the geometry strategy does not always rely on the user to provide control points. Instead, according to the learning agent's current knowledge, radials on which the appearance strategy category has a weighting large enough (i.e., α above a pre-defined threshold) use only the appearance category to localize the contour points, and these points becomes control points for the spline. Secondly, as the user examples directly provide positions as control points, f_G depends on the positions of the user examples (by contrast all f_{A_i} are independent from the user input). Therefore, to perform policy gradient in such a dependent situation, an iterative flow is used: after the weightings on a new example radial are learned, they are used together with the example's position to update weightings of every previous example radial in the same image; then the updated weightings of previous example radials are used to update the weightings of the latest example. The iterative procedure continues until the weightings on example radials remain unchanged.

With the two-layer reinforcement learning, optimal weightings are learnt for radials in the appearance feature space where examples are available from the user. For segmenting an image, weightings on all radials need to be estimated based on the learned knowledge. The estimation is realized through Kriging [10], which interpolates the value of a random field.

2.2 Validation and Data Collection

The MR LV data used for validation were from a database of 56 subjects acquired from a 1.5T MR scanner (Siemens Sonata 1.5T, Erlangen, Germany) and a trueFISP sequence (in-plane pixel resolution = 1.1 - 2 mm, slice thickness = 7 mm). For deriving the ground truth data, manual delineation was performed by two expert clinicians using CMRtools (Cardiovascular Imaging Solutions Ltd., London, UK).

3 Results

While Fig. 1 shows some examples of applying the proposed method to different structures, detailed validation is carried out on 3D LV endocardial borders of the end-diastolic frame and the end-systolic frames for the 56 subjects. For segmenting each slice, the user is required to provide three anchor points to initialize the algorithm: the centre of the LV blood pool and the RV/LV junctions on the endocardial border. The RV/LV junctions and the points for corrections provided by the user are both used as training examples. The candidate strategies used in the appearance category include the maximal derivative along the radial, a 1D intensity profile and a 1D gradient profile, as they are popularly used in detecting endocardial borders. The strategies as well as the parameterization used are intended as an example for the reinforcement learning framework. For different applications, the reader may use other specific features, searching and parameterization techniques by switching to the corresponding state-action space while maintaining the general reinforcement learning framework.

Table 1. The accuracy (in *mm*) and the user input required by the proposed method and manual segmentation using CMRtools for segmenting the end-diastolic (ED) and the end-systolic (ES) frames (The step size is 0.2 for gradient calculation, $\gamma = 1$ and the threshold of α is set to 0.9)

	Proposed Method		Manual Segmentation		Improvement
Frame & Location	Error	Avg. No. of User Points	Inter-Observer Variability of Accuracy	Avg. No. of User Points	User Interaction Reduction
ED-Base	0.81±0.25	3.1	1.15±0.27	7.0	55.7%
ED-Mid	0.86±0.24	3.5	1.00±0.33	6.9	49.3%
ED-Apex	0.70±.23	3.1	1.52±0.73	6.6	53.0%
ED-All	0.81±0.14	19.6	1.18±0.34	40.9	52.1%
ES-Base	0.67±0.22	2.7	1.23±0.47	5.4	50.0%
ES-Mid	0.74±0.25	3.1	1.08±0.59	5.9	47.5%
ES-Apex	0.53±0.20	2.7	1.49±0.63	5.1	47.1%
ES-All	0.67±0.17	17.5	1.27±0.45	33.1	47.1%

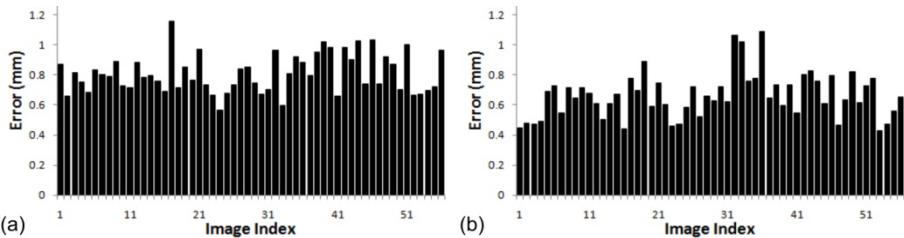


Fig. 2. Detailed segmentation results for the end-diastolic (a) and the end-systolic frame (b)

The segmentation result is compared with the manual segmentation for both accuracy and the number of user input points required, as summarized in Table 1. It is evident that for both the end-diastolic and the end-systolic frames, the amount of user interaction is significantly reduced whilst maintaining the overall accuracy as shown in Fig. 2. The paired t-test shows that the improvement is significant for both frames

with $p \ll 0.0001$. It is worth noting that the result of the proposed method has an error much smaller than the inter-observer variability. The reason is that in cardiac MR, segmentation is not just dependent on geometrical features or appearance information (e.g. dealing with papillary muscle), but also other prior knowledge, which leads to higher inter-observer variability. The ability through reinforcement learning to acquire this specific knowledge minimizes the error involved, and thus gives improved consistency. More error metrics are summarized in Table 2.

Table 2. The Jaccard index and the maximal radial difference of the proposed method

Location	End-Diastole		End-Systole	
	Jaccard Index	Max. Radial Diff.	Jaccard Index	Max. Radial Diff.
Basal	0.933±0.022	2.55 mm	0.915±0.031	2.41 mm
Middle	0.928±0.019	2.60 mm	0.906±0.036	2.49 mm
Apical	0.926±0.025	2.04 mm	0.889±0.048	2.00 mm
All	0.929±0.012	2.60 mm	0.905±0.030	2.49 mm

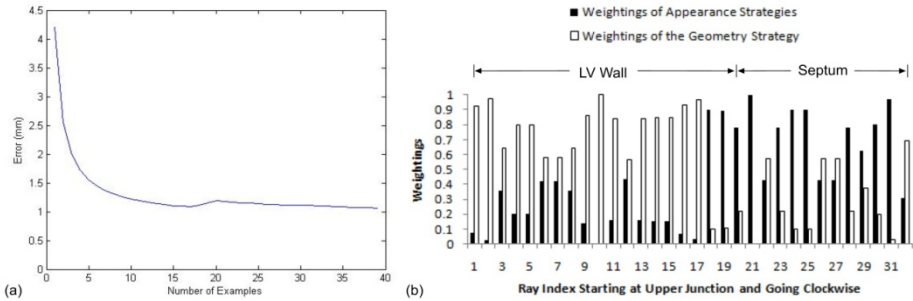


Fig. 3. (a) Learning curve; (b) Learned cross-category weightings for radials in a slice

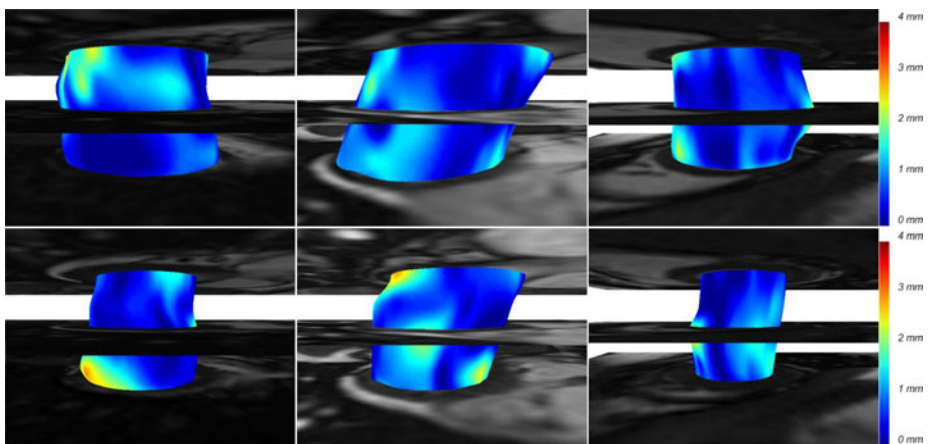


Fig. 4. Three instances of the end-diastolic (top) and end-systolic (bottom) frames showing surface localization errors. High accuracy is achieved with the proposed algorithm.

Further analysis is carried out to evaluate the reinforcement learning process. A fixed data set of 20 3D cardiac MR images at the end-systolic frame is used for testing the learning behavior, while the rest of the data are used to interact with the user as in the experiment above. By using the proposed method, the model is updated whenever a new user example is given. The updated model is then used to segment all the testing images without correction. The results are shown in Fig. 3 (a). It is clear that with an increasing number of examples, the model is able to perform the segmentation more accurately. Fig. 3 (b) shows the cross-type weightings derived from the model for all the points around the endocardial border for a mid-ventricular slice at end-systole. As expected, it can be seen that at the LV wall region, the segmentation relies more on the geometry strategy, while at the septal region it relies more on the appearance strategies. Fig. 4 illustrates three examples of the diastolic and systolic volumes of the LV, highlighting regional errors involved in the segmentation results.

4 Conclusion

In this paper, we have presented a general semi-automatic scheme that learns from the user behavior to perform medical image segmentation. The algorithm determines the weightings of different segmentation strategies adaptively, which otherwise need to be either empirically defined or obtained from off-line training, and thus cause overfitting or poor generalization. The adaptation is achieved by a two-layer reinforcement learning algorithm that learns from the user examples. Detailed validation on MR LV segmentation demonstrates the potential clinical value of the algorithm.

References

1. Xu, C., Prince, J.L.: Snakes, shapes, and gradient vector flow. *IEEE Transactions on Image Processing* 7, 359–369 (1998)
2. Chan, T., Vese, L.: Active contour without edges. *IEEE Transactions on Image Processing* 10, 266–277 (2001)
3. Cootes, T.F., Hill, A., Taylor, C.J., Haslam, J.: The use of active shape models for locating structures in medical images. In: Barrett, H.H., Gmitro, A.F. (eds.) *IPMI 1993*. LNCS, vol. 687, pp. 33–47. Springer, Heidelberg (1993)
4. Wang, Y., Staib, L.H.: Boundary finding with prior shape and smoothness models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22, 738–743 (2000)
5. Davatzikos, C., Tao, X., Shen, D.: Hierarchical active shape models, using the wavelet transform. *IEEE Transactions on Medical Imaging* 22, 414–423 (2003)
6. Lekadir, K., Keenan, N.G., Pennell, D.J., Yang, G.-Z.: An inter-landmark approach to 4-D shape extraction and interpretation: application to myocardial motion assessment in MRI. *IEEE Transactions on Medical Imaging* 30, 52–68 (2011)
7. Sutton, R.S., Barto, A.G.: *Reinforcement learning: an introduction*. The MIT Press, Cambridge (1998)
8. Watkins, C.J.C.H.: *Learning from delayed rewards*. PhD Thesis, Cambridge University (1989)
9. Akima, H.: *A method of smooth curve fitting* (1969)
10. Matheron, G.: *Principles of geostatistics*. *Economic Geology* 58, 1246–1266 (1963)

Robust Deformable-Surface-Based Skull-Stripping for Large-Scale Studies

Yaping Wang^{1,2}, Jingxin Nie², Pew-Thian Yap², Feng Shi², Lei Guo¹,
and Dinggang Shen^{2,*}

¹ School of Automation, Northwestern Polytechnical University,
Xi'an, Shaanxi Province, China

² Department of Radiology and BRIC, University of North Carolina at Chapel Hill, USA
dgshen@med.unc.edu

Abstract. Skull-stripping refers to the separation of brain tissue from non-brain tissue, such as the scalp, skull, and dura. In large-scale studies involving a significant number of subjects, a fully automatic method is highly desirable, since manual skull-stripping requires tremendous human effort and can be inconsistent even after sufficient training. We propose in this paper a robust and effective method that is capable of skull-stripping a large number of images accurately with minimal dependence on the parameter setting. The key of our method involves an initial skull-stripping by co-registration of an atlas, followed by a refinement phase with a surface deformation scheme that is guided by prior information obtained from a set of real brain images. Evaluation based on a total of 831 images, consisting of normal controls (NC) and patients with mild cognitive impairment (MCI) or Alzheimer's Disease (AD), from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database indicates that our method performs favorably at a consistent overall overlap rate of approximately 98% when compared with expert results. The software package will be made available to the public to facilitate neuroimaging studies.

1 Introduction

Automated skull-stripping is an indispensable tool for large-scale multi-site studies, such as the Alzheimer's Disease Neuroimaging Initiative (ADNI) [1], where thousands of MRI scans need to be processed. An accurate and highly automated algorithm is hence highly desirable, preferably avoiding any form of human intervention that can possibly cause bias and inter-expert differences. The inaccuracy of skull-stripping may result in error in cortical thickness estimation, volumetric analysis, brain atrophy estimation, and registration between structural and functional images (fMRI and PET). For instance, the unremoved dura can result in an overestimation of cortical thickness [2], whereas the unintended removal of brain tissue may lead to underestimation of cortical atrophy. These errors cannot be recovered in subsequent processing stages. The existing skull-stripping methods can generally be divided into three categories: 1) morphology-based methods, 2) surface-based methods, and 3) meta methods.

* Corresponding author.

Morphology-based methods extract the brain through a series of thresholding and morphological operations [3, 4]. In [4] for example, a heuristic approach utilizing thresholding, morphological opening, and connected component analysis is devised to deal with specific problems encountered in skull-stripping. Another popular method is the Brain Surface Extractor (BSE) [5], which removes non-brain tissue using a combination of anisotropic diffusion filtering, edge detection, and mathematical morphology. Also, a 3D watershed transform, modified by combining pre-flooding to avoid over-segmentation, is proposed in [6]. A graph cuts based method is introduced in [7], utilizing intensity thresholding followed by removal of narrow connection using a graph theoretic image segmentation technique. *While some published approaches are effective, morphology-based methods generally involve some degree of user interaction, and are also sensitive to the scanning parameters and intensity inhomogeneity.*

Surface-based methods rely mainly on the image gradient information to locate the brain surface, which is modeled by an active contour. In [8], a tessellated ellipsoidal template is deformed to the inner surface of the skull, and is then iteratively deformed using forces derived from image gradients and a curvature constraint. Brain Extraction Tool (BET) [9] uses a deformable model that evolves a surface to fit the brain boundary by application of a set of locally adaptive forces. These forces account for surface smoothness and voxel intensity changes in the surface vicinity. Hybrid Watershed Algorithm (HWA) [10] combines watershed algorithm developed in [6] with a deformable-surface model, using surface curvature and statistics of the distance of the surface to the center of gravity (COG) to detect and correct inaccuracies in brain extraction. The above-described methods are generally more robust and less sensitive to image artifacts, and require less human interaction. *However, these methods require that the initial active contour should be close enough to the targeted surface to avoid local minima; and failure to meet this requirement can cause unpredictable results.*

Meta methods combine various skull-stripping methods for achieving the best outcome [11, 12]. However, they rely heavily on the component methods and do not solve the fundamental problems pertaining to these methods.

Due to the presence of imaging artifacts, anatomical variability, and varying contrast properties, most of these techniques, however, do not give satisfactory results over a wide range of scan types and neuroanatomies without some form of manual intervention. In this paper, we focus on a surface-based method, since it gives us the flexibility of incorporating both surface geometry and atlas-based prior information into the skull-stripping process. For a good initial extraction of the brain, we utilize FLIRT [13] and Demons-based registration [14] to warp an atlas, and hence its brain mask, to the targeted image for initial skull-stripping. We then employ a probability map, incorporating prior information gathered from a population of real brain images with manual extractions, to guide surface evolution for refinement of the skull-stripping result. This allows us to guide the skull-stripping process using realistic shape and topological constraints to weight the forces in driving surface deformation, thus giving us more accurate and consistent results over a large number of images.

2 Method

The skull-stripping method consists of the following steps: 1) Initialization by warping an atlas to the target image, where the brain mask for the atlas is used for

approximate skull-stripping, and 2) Refinement of the former result with a deformable-surface, guided by a brain probability map generated from a set of real images. The details are given in the following.

2.1 Initialization

A good initialization of the deformable surface is important for avoiding local minima and suboptimal solutions. To facilitate more accurate parameter estimation for better positioning of the initial active contour, FLIRT [13] and the Demons registration algorithm [14] are employed to help mask out non-brain voxels. Demons registration works especially well in the image regions with clear intensity changes. In a T1-weighted MR image, the brain boundary appears as a surface with low cerebrospinal fluid (CSF) level intensity, compared with the GM and the skull, thus providing good contrast for alignment. We first register the ICBM template (with skull) [15] to the target image and then use the accompanying brain mask, which is dilated for avoiding over-skull-stripping, as an initial skull-stripping, which will be further refined using a deformable-surface-based approach as described below.

2.2 Deformable-Surface-Based Skull-Stripping

2.2.1 Brain Probability Map

A probability map, indicating the likelihood of whether a voxel belongs to a part of the brain tissue, is obtained from a set of manually skull-stripped images. Specifically, we utilize the ICBM template as the reference image, onto which a set of images (with skulls) are aligned first linearly (FLIRT) [13] and then non-linearly (Demons) [14]. The manually delineated brain masks of the images are warped to the template space using the same set of transformations. From the aligned brain masks, we can evaluate for each voxel the probability that it belongs to the brain by simply computing the fraction of brain masks that consider this voxel a part of the brain. That is, for each point (i, j, k) , the probability is $p(i, j, k) = \sum_{n=1}^N B_n(i, j, k) / N$, where B_n represents the n -th brain mask out of a total of N . From Fig. 1, we can see that the probability map has a value of 1 for the majority of the brain, except the boundary where the voxel memberships are ambiguous. This probability map is warped onto the target image to guide skull-stripping using a deformable surface. The probability map helps greatly restrict the extent of deformation of the surface, thus minimizing the chance of it falling in a less desirable region.



Fig. 1. Brain probability map overlaid on the brain image, shown in three different views

2.2.2 Surface Deformation

Following the work of Smith [9], the implementation¹ of the evolution equation is accomplished with a parametric active contour. The brain boundary is modeled by a surface tessellated using connected triangles. The surface deformation is driven by three main forces: 1) force from the smoothness constraint, 2) force from the intensity gradient in the surface vicinity, and 3) force from the brain probability map.

The force pertaining to the smoothness constraint consists of two parts:

$$\mathbf{F}_1 = c_1 \mathbf{u}_t + c_2 \mathbf{u}_n, \quad (1)$$

where \mathbf{u}_t and \mathbf{u}_n are the tangential and normal components of \mathbf{u} . \mathbf{u} is the difference vector between the positions of the current vertex and its mean neighboring vertices, i.e.,

$$\mathbf{u} = \frac{1}{M} \sum_{m=1}^M \mathbf{x}_m - \mathbf{x}_0, \quad \mathbf{u}_n = (\mathbf{u} \cdot \mathbf{n})\mathbf{n}, \quad \mathbf{u}_t = \mathbf{u} - \mathbf{u}_n, \quad (2)$$

where \mathbf{n} is the surface outward normal at current vertex \mathbf{x}_0 , M is the number of neighboring vertices, and \mathbf{x}_m is the m -th neighboring vertex. The sole role of the tangent force \mathbf{u}_t is to keep all vertices in the surface equally spaced, i.e., moving them only within the surface. The normal force \mathbf{u}_n is a curvature-reducing force to ensure smoothness of the surface during the evolution process. Usually c_1 is set as 0.5. To ensure that the surface is sufficiently smooth and meanwhile avoids underestimation of curvature at parts of the surface, c_2 is a nonlinear function adaptive to local surface geometry. For this purpose, we need first to determine the local radius of curvature r : $r = (d^2)/(2|\mathbf{u}_n|)$, where d is the mean distance from vertex to neighboring vertices over the whole surface. Weight c_2 is then defined using a sigmoid function:

$$c_2 = 0.5(1 + \tanh(l_2 * (1/r - l_1))) \quad (3)$$

with $l_1 = 0.5(1/r_{MIN} + 1/r_{MAX})$ and $l_2 = 6/(1/r_{MIN} - 1/r_{MAX})$. Here, the values for r_{MIN} and r_{MAX} are empirically optimized for typical geometries in the human brain and are set as 3.33 mm and 10 mm, respectively. High local mean surface curvature is hence penalized to achieve surface smoothing.

The second force is derived from the image intensity information:

$$\mathbf{F}_2 = \frac{2(I_{MIN} - G_1)}{I_{MAX} - G_2} \mathbf{n}, \quad (4)$$

where I_{MIN} and I_{MAX} are the minimum and maximum of local intensities quantified along the normal direction pointing inward from the current vertex:

$$I_{MIN} = \max(G_2, \min(G_m, I(0), I(1), \dots, I(S_1)));$$

$$I_{MAX} = \min(G_m, \max(G_s, I(0), I(1), \dots, I(S_2))).$$

The search range pertaining to the minimum and maximum intensities are S_1 and S_2 , respectively. Typically, S_1 is set as 20 mm and $S_2 = S_1/2$. Here the image intensity minimum G_2 , the median intensity G_m , and the threshold G_s that separates brain and non-brain matter are approximately estimated according to the cumulative intensity

¹ The proposed method is a modification of the original BET algorithm.

histogram of the brain image. They're used to avoid voxels that are too dark or too bright. G_1 in Equation (4) is a locally estimated intensity threshold which is used to distinguish between brain and non-brain:

$$G_1 = (I_{MAX} - G_2) * f + G_2. \quad (5)$$

Parameter f is called the fractional intensity threshold and falls between the range of 0 and 1.

The third force is derived from the brain probability map obtained in **Section 2.2.1**:

$$\mathbf{F}_3 = (p - 0.5)\mathbf{n}, \quad (6)$$

where $p \in [0,1]$ is the probability value from the map. At a point where the surface is close to the brain boundary, $p \approx 0.5$, and \mathbf{F}_3 is small; when $p > 0.5$, the point is likely within the brain, and an outward force is exerted; when $p < 0.5$, an inward force is exerted. This restricts the searching of the brain boundary in this range and drives it to find the true boundary. \mathbf{F}_3 is directly added to \mathbf{F}_2 with a certain weight c_4 , which we set as 1 here. Both of them act in the direction of \mathbf{n} . Thus, for each vertex i at iteration t , the update equation is:

$$\mathbf{x}_i^{t+1} = \mathbf{x}_i^t + [\mathbf{F}_1 + c_3(\mathbf{F}_2 + c_4\mathbf{F}_3)], \quad (7)$$

where $c_3 = 0.05 * d$.

3 Experimental Results

3.1 Dataset

The proposed algorithm was evaluated based on the ADNI dataset. Subjects were recruited from over 50 sites across the U.S. and Canada. MRI scans were acquired using GE, Philips, and Siemens scanners. The baseline T1-weighted MRI scans of 831 subjects (230 NC, 403 MCI, 198 AD; ages range: 55~90 years) were used in this work. We randomly selected images of 25 subjects from each group for generating the respective probability maps; the rest of the images were used as testing data for evaluation of all considered methods. For consistency, all images were resampled to be of dimensions $256 \times 256 \times 256$ and resolution $1 \times 1 \times 1 \text{ mm}^3$. Nonparametric nonuniform intensity normalization (N3) [16] was performed for correcting the intensity inhomogeneity.

3.2 Qualitative Evaluation

The following 5 methods were evaluated: 1) BET, 2) BET-B: BET with bias field correction and neck cleanup (option “-B”), 3) BSE, 4) HWA, and 5) the proposed method. Let P and Q represent the extracted brain image and the manually skull-stripped image, respectively. We measured their similarity using the Dice ratio: $D(P, Q) = 2|P \cap Q| / (|P| + |Q|)$. For each method compared, except our method, we determined for each image the result that gave the best Dice ratio by grid search over a range of parameter values, compared with manual skull-stripping results by an expert. The following is the range of values considered for the parameters (in

MATLAB notation): BET, fractional intensity threshold = [0.1:0.05:0.8]; BET-B, fractional intensity threshold = [0.1:0.05:0.8], option “-B” turned on; BSE, diffusion constant = [5:5:60], edge detection constant = [0.3:0.05:0.8], diffusion iterations = [3:1:5]; HWA, the default parameters with or without “-less”, “-more”, “-atlas”, “-less -atlas” or “-more -atlas” were used. For the proposed method, a fractional intensity threshold value of 0.6 was used for all images. Typical results given by these 5 methods are shown in Fig. 2. For the image of Fig. 2, the best parameters selected are listed below: BET, fractional intensity threshold = 0.75; BET-B, fractional intensity threshold = 0.4, option “-B” turned on; BSE, diffusion constant = 5, edge detection constant = 0.7, diffusion iterations = 3; HWA, the default parameters were used; the proposed method, fractional intensity threshold = 0.6. The arrows in Fig. 2 highlight several problematic areas. BET typically over skull-strips the anterior frontal cortex, anterior temporal cortex, posterior occipital cortex, and the cerebellar areas. BSE typically leaves some dura unremoved in these regions. HWA retains most of the brain tissue, but leaves some dura intact. The proposed method successfully overcomes all these problems.

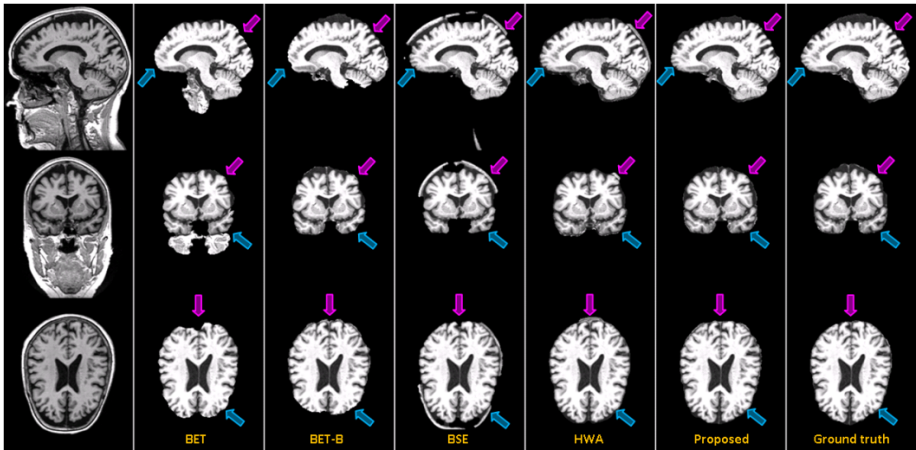


Fig. 2. Typical skull-stripping results of existing methods, compared with the proposed method

3.3 Quantitative Evaluation

Similar to the qualitative evaluation, we compared the proposed method with BET, BET-B, BSE and HWA. And for each method compared, except our method, the optimal parameters for each image were selected as described in Section 3.2. For the proposed method, a fractional intensity threshold value of 0.6 was used for all images as described in Section 3.2. As can be appreciated from Fig. 3, the proposed method yields consistently the best results when compared with all other methods, despite the fact that one single set of parameters was used for all images, whereas optimized parameters were used for other methods. Fig. 4 shows that the performance of the proposed method is insensitive to different fractional intensity threshold values. Of note, on a 2.8 GHz machine, the proposed method took approximately 2 minutes for

registration using FLIRT and Demons, and an extra 1 minute for skull-stripping an image. Other methods typically took 60-70 seconds, except for BET when used with option “-B”, which typically took 20-30 minutes.

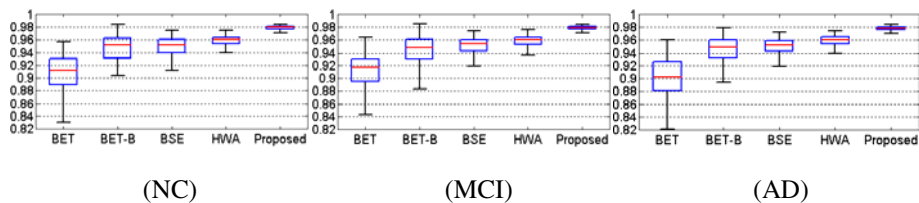


Fig. 3. Distributions of Dice ratios for different methods and different subject groups

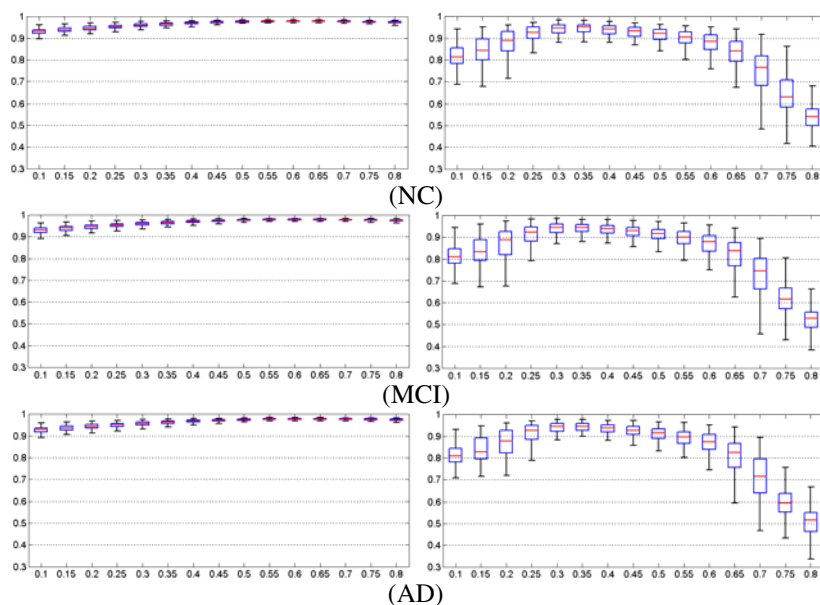


Fig. 4. Distributions of Dice ratios over different fractional intensity threshold values for the proposed method (left) and BET-B (right). The proposed method is robust to parameter variation.

4 Conclusion

We have proposed in this paper a robust and effective skull-stripping method that is capable of skull-stripping a large number of images automatically without human intervention. Our method combines an atlas-based approach, for coarse skull-stripping, with a deformable-surface-based approach, for more localized refinement. Experimental results indicate that our method yields highly consistent results using

only a set of fixed parameters. The software package will be released to the public for facilitating neuroimaging studies.

Acknowledgments. This work was supported in part by NIH grants EB008374, EB006733, and EB009634.

References

1. Mueller, S.G., Weiner, M., Thal, L.J., Petersen, R.C., Jack, C., Jagust, W., Trojanowski, J.Q., Toga, A.W., Beckett, L.: The Alzheimer's disease neuroimaging initiative. *Neuroimaging Clin. N. Am.* 15, 869–877 (2005)
2. van der Kouwe, A.J., Benner, T., Salat, D.H., Fischl, B.: Brain morphometry with multiecho MPRAGE. *NeuroImage* 40, 559–569 (2008)
3. Höhne, K.H., Hanson, W.A.: Interactive 3D segmentation of MRI and CT volumes using morphological operations. *J. Comput. Assist. Tomogr.* 16, 285–294 (1992)
4. Lemieux, L., Hagemann, G., Krakow, K., Woermann, F.G.: Fast, accurate, and reproducible automatic segmentation of the brain in T1-weighted volume MRI data. *Magn. Reson. Med.* 42, 127–135 (1999)
5. Shattuck, D.W., Sandor-Leahy, S., Schaper, K.A., Rottenberg, D.A., Leahy, R.M.: Magnetic resonance image tissue classification using a partial volume model. *NeuroImage* 13, 856–876 (2001)
6. Hahn, H.K., Peitgen, H.-O.: The Skull Stripping Problem in MRI Solved by a Single 3D Watershed Transform. In: Delp, S.L., DiGoia, A.M., Jaramaz, B. (eds.) *MICCAI 2000*. LNCS, vol. 1935, pp. 134–143. Springer, Heidelberg (2000)
7. Sadananthan, S.A., Zheng, W., Chee, W.L., Zagorodnov, V.: Skull stripping using graph cuts. *NeuroImage* 49, 225–239 (2010)
8. Dale, A.M., Fischl, B., Sereno, M.I.: Cortical surface-based analysis. I. Segmentation and surface reconstruction. *NeuroImage* 9, 179–194 (1999)
9. Smith, S.M.: Fast robust automated brain extraction. *Hum. Brain Mapp.* 17, 143–155 (2002)
10. Ségonne, F., Dale, A.M., Busa, E., Glessner, M., Salat, D., Hahn, H.K., Fischl, B.: A hybrid approach to the skull stripping problem in MRI. *NeuroImage* 22, 1060–1075 (2004)
11. Rex, D.E., Shattuck, D.W., Woods, R.P., Narr, K.L., Luders, E., Rehm, K., Stoltzner, S.E., Rottenberg, D.A., Toga, A.W.: A meta-algorithm for brain extraction in MRI. *NeuroImage* 23, 625–637 (2004)
12. Leung, K.K., Barnes, J., Modat, M., Ridgway, G.R., Bartlett, J.W., Fox, N.C., Ourselin, S., ADNI: Brain MAPS: An automated, accurate and robust brain extraction technique using a template library. *NeuroImage* 55, 1091–1108 (2011)
13. Jenkinson, M., Smith, S.: A global optimisation method for robust affine registration of brain images. *Medical Image Analysis* 5, 143–156 (2001)
14. Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Non-parametric Diffeomorphic Image Registration with the Demons Algorithm. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) *MICCAI 2007, Part II*. LNCS, vol. 4792, pp. 319–326. Springer, Heidelberg (2007)
15. Holmes, C.J., Hoge, R., Collins, L., Woods, R., Toga, A.W., Evans, A.C.: Enhancement of MR images using registration for signal averaging. *J. Comput. Assist. Tomogr.* 22, 324–333 (1998)
16. Sled, J.G., Zijdenbos, A.P., Evans, A.C.: A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Trans. Med. Imaging* 17, 87–97 (1998)

Confidence-Guided Sequential Label Fusion for Multi-atlas Based Segmentation

Daoqiang Zhang^{1,2}, Guorong Wu¹, Hongjun Jia¹, and Dinggang Shen¹

¹ Dept. of Radiology and BRIC, University of North Carolina at Chapel Hill, NC 27599

² Dept. of Computer Science and Engineering, Nanjing

University of Aeronautics and Astronautics, Nanjing 210016, China

{zhangd, grwu, jiahj, dgshen}@med.unc.edu

Abstract. Label fusion is a key step in multi-atlas based segmentation, which combines labels from multiple atlases to make the final decision. However, most of the current label fusion methods consider each voxel equally and independently during label fusion. In our point of view, however, different voxels act different roles in the way that some voxels might have much higher confidence in label determination than others, i.e., because of their better alignment across all registered atlases. In light of this, we propose a sequential label fusion framework for multi-atlas based image segmentation by hierarchically using the voxels with high confidence to guide the labeling procedure of other challenging voxels (whose registration results among deformed atlases are not good enough) to afford more accurate label fusion. Specifically, we first measure the corresponding labeling confidence for each voxel based on the k -nearest-neighbor rule, and then perform label fusion sequentially according to the estimated labeling confidence on each voxel. In particular, for each label fusion process, we use not only the propagated labels from atlases, but also the estimated labels from the neighboring voxels with higher labeling confidence. We demonstrate the advantage of our method by deploying it to the two popular label fusion algorithms, i.e., majority voting and local weighted voting. Experimental results show that our sequential label fusion method can consistently improve the performance of both algorithms in terms of segmentation/labeling accuracy.

1 Introduction

In computational anatomy, the accurate image segmentation and labeling is a critical step for many clinical studies, such as pathology detection and brain parcellation. Although a lot of automatic image segmentation methods have been investigated, it is still a hot topic in medical image analysis.

Recently, multi-atlas based segmentation methods have shown great success in segmenting brain into anatomical structures [1-7]. A typical multi-atlas based segmentation procedure contains two major steps: 1) image registration step to register each atlas image to the target image and warp the corresponding label image by following the same estimated deformation field; and 2) label fusion step to combine the multiple propagated labels from different atlases to obtain the final labels

of the target image by some heuristics. The current methods usually complete this procedure by independently performing the two steps, without considering the relationship between them.

A number of label fusion strategies have been proposed for multi-atlas based segmentation in the literature. Among them, majority voting (MV) is probably the simplest one and has been widely used in medical image segmentation. In MV, the candidate segmentations from each atlas are equally weighted and the label with largest agreement from all atlases is assigned as the final label. A natural extension of MV is to improve from simple averaging to adaptive weighted averaging. In [4], various weighting strategies are categorized into two groups, i.e., global weighted voting and local weighted voting, and it has been shown that the local weighted method outperforms the global solution when segmenting high-contrast brain structures. On the other hand, instead of using all atlases, using a selected subset of atlases usually results in improved performance [5]. The recent trend is to use the more advanced learning based methods to further improve performances of label fusion. In [6], a probabilistic label fusion method is proposed to explicitly model the relationship between the atlas and the target image. To avoid the possible registration error, a non-local label fusion method based on the patch-based strategy is proposed in [7], which is widely used in the machine learning and computer vision community.

One common limitation of existing methods is that the label fusion procedure treats each voxel independently and equally. Although it simplifies the labeling procedure, the coherent spatial correlations between neighboring voxels are ignored, which are actually very useful to achieve the accurate and robust segmentation. Moreover, due to the huge anatomical variations in the population, the registration accuracy varies not only across different subjects but also at different locations of the same subject. As a result, some voxels may have more reliable estimation on labels because of the more accurate alignment. In light of this, we should first perform label fusion on voxels with higher labeling confidence, and then use the estimated (usually more reliable) labels to guide the label fusion of the neighboring voxels with lower labeling confidence, in a sequential way.

Based on this observation, in this paper, we present a novel sequential label fusion framework for multi-atlas based segmentation. First, a novel criterion on labeling confidence is defined which considers not only the similarity but also the matching consistency of the two underlying local patches. Base on the labeling confidence on each voxel, the label procedure is sequentially performed from the voxels with high confidence to the voxels with low confidences. Thus, our method contains two subsequent steps, i.e., 1) labeling confidence estimation, and 2) sequential label fusion. Specifically, we first estimate the labeling confidence for each voxel based on the k -nearest neighbor (k -NN) rule [8], and then perform label fusion sequentially according to the measured confidence. When labeling each voxel, we use not only the propagated labels from the atlases, but also the already estimated labels from the neighboring voxels with higher labeling confidence. To the best of our knowledge, the proposed idea on sequential label fusion is new in the multi-atlas based segmentation. Our framework is general and can be integrated with any other label fusion methods. As confirmed in the experiment on NIREP dataset [9], our method is able to significantly improve the labeling accuracy, in comparison with the two popular label fusion algorithms: majority voting (MV) and local weighted voting (LWV).

2 Methods

In this section, we will present the sequential label fusion framework which consists of two steps: 1) labeling confidence estimation, and 2) sequential label fusion. We will first introduce how to estimate the registration confidence through k -NN searching in Section 2.1, and then detail the flow chart of the sequential label fusion framework in Section 2.2.

2.1 Labeling Confidence Estimation

Here, we use T to denote the target image to be labeled, and assume that all atlases and their associated label images have been already registered with the target image T , represented as $\{(A_i, L_i) | i=1, \dots, M\}$. Thus, the target image T and all atlases are considered in the same domain Ω , and the procedure of label fusion on each voxel p in target image T is performed among a stack of $\{A_i(p), L_i(p)\}$ for each voxel p ($p \in \Omega$).

To measure the importance of each voxel in label fusion, we propose a novel criterion, called *labeling confidence*, which is based on local image appearance and matching consistency. Given voxel p in T , the calculation of its labeling confidence with respect to atlas A_i is performed in two steps:

(1) *Forward Matching*: We first compute the patch difference between *fixed voxel* p in T and *moving voxel* o in the forward search neighborhood (denoted as $N_f(p)$ and displayed as the red circle in Fig. 1) in atlas A_i , and then perform a k -NN search to find voxel p 's k nearest neighbors q_s^i (displayed as blue squares in the atlases, Fig.1). We denote $Q_i(p) = \{q_s^i | s=1, \dots, k\}$ as the set of those k neighbors.

(2) *Backward Matching*: For each voxel q_s^i in $Q_i(p)$ (p), we perform another backward matching as follows. We first compute the patch difference between fixed voxel o and the moving voxels p' in the backward search neighborhood (denoted as $N_b(q_s^i)$ and displayed as the blue circles in Fig. 1) in the T , and then perform a k -NN search to find voxel q_s^i 's k nearest neighbors $u_s^i(t)$ (displayed as green triangles in T in Fig.1). Similarly, we denote $P_s^i(p) = \{u_s^i(t) | t=1, \dots, k\}$ as the set of those k neighbors.

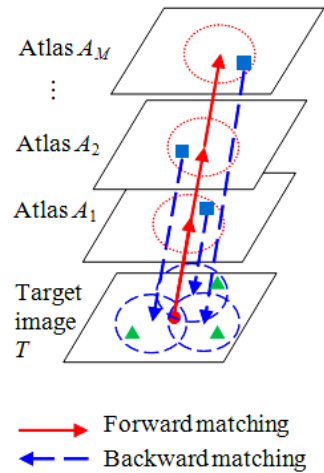


Fig. 1. Illustration on forward and backward matching for labeling confidence estimation with $k=1$

Then, the labeling confidence of the voxel p w.r.t. atlas A_i is computed as:

$$C_i(p) = \frac{1}{k^2} \sum_{s=1}^k \sum_{t=1}^k \exp(-\delta_1^{-1}(p)D(p, q_s^i)) \exp(-\delta_2^{-1}(p)\|p - u_s^i(t)\|_2^2), \quad (1)$$

with

$$\delta_1(p) = \min_{q_s^i \in Q_i(p)} D(p, q_s^i) + \varepsilon, \text{ and } \delta_2(p) = \min_{u_s^i(t) \in P_i^t(p)} \|p - u_s^i(t)\|_2^2 + \varepsilon$$

used for normalization, where ε is a small constant. In Eq. 1, the first term after the sum symbols is the normalized patch difference (based on local image appearance) under the distance function D , while the last term measures the matching consistency. Here, the intuition is that for good registration the distance between p and $u_s^i(t)$, weighted by the local appearance similarity, are required to be as small as possible after forward and backward matching. In this way, both image appearance and the consistency are embedding in the measurement of labeling confidence.

The confidence degree estimated in Eq. 1 for each atlas A_i will be used as a weighting map for weighted voting in the following section. Moreover, we can average the confidence degree $C_i(p)$ from individual atlas A_i to measure overall confidence degree at each target image voxel p , i.e., $C(p) = (1/M) \sum_{i=1}^M C_i(p)$, which will be used to guide the sequential label fusion in next subsection. After that, we call $C(p)$ as *labeling confidence*.

2.2 Sequential Label Fusion

The estimated registration confidence map $C = \{C(p) | p \in \Omega\}$ reflects the registration confidences of different voxels in the target image T . Thus, based on the confidence map, we present a sequential label fusion framework, as shown in Fig. 2. For simplicity, we focus on the binary segmentation of anatomical structures, with label ‘1’ indicating the structure of interest and ‘0’ for the others. We first generate the initial voxels set V for the target image T according to the union of all L_i s. Then, we start from the voxel p in V with the highest labeling confidence of $C(p)$ and compute its soft label (i.e., segmentation probability) by

$$L(p) = \frac{1}{Z_p} \left[l(p | L_1, \dots, L_M) + \sum_{p' \in NH(p)} L(p') \exp(-\frac{1}{\sqrt{3}r} \|p - p'\|_2^2) \right]. \quad (2)$$

Here, $l(p | L_1, \dots, L_M)$ denotes the propagated soft label from atlases for voxel p which can be obtained by using any label fusion algorithms, $NH(p)$ is the set of the neighboring voxels in the neighborhood $N(p)$ which satisfies $C(p') > C(p)$, and Z_p is a normalizing term to make $L(p) \in [0, 1]$. It is obvious that the procedure of label fusion on the voxel of lower confidence will be guided by the neighboring voxels which have higher confidence and are thus able to correctly identify their labels with low risk (see the second term in the square bracket in Eq. 2). This process is iterated until all voxels in the volume have their labels determined. Finally, the final label of voxel p is computed as: $\text{sign}(L(p) - 0.5)$.

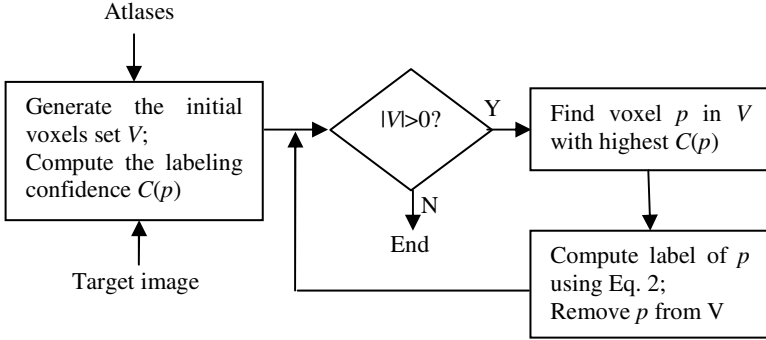


Fig. 2. Flow chart of the proposed sequential label fusion framework

3 Experiments

In this section, we evaluate the performance of the proposed sequential label fusion method. We integrate our sequential label fusion framework with the two most widely used algorithms, i.e., majority voting (MV) and local weighted voting (LWV). In LWV, the weights are based on the mean square distance on the local regions ($r \times r \times r$) between target image and atlas [4]. We denote the new MV and LWV algorithms equipped with our sequential label fusion as SC-MV and SC-LWV, respectively. In order to show the advantage of sequential labeling strategy, we also compare with two other variants which only use the confidence maps C_i ($i=1, \dots, M$) as weights but still treat each voxel independently. We denote these two methods as C-MV and C-LWV, respectively. In the following experiment, we perform all these six algorithms on the NIREP database [9].

The NIREP database [9] consists of 16 T1-weighted MR image (8 normal males and 8 females) with 32 manually delineated regions of interest (ROIs). For each of the ROIs, a Leave-One-Out (LOO) cross-validation is performed to test the segmentation performance, and the averaged Dice overlap measures are reported. Specifically, at each LOO fold, all other 15 subjects are used as the atlases and aligned onto the remaining image (used as target image) for guiding the segmentation. In our method, the sizes of forward and backward search neighborhood N_f and N_b are set to 5, and a local patch of size 3 is used to compute the Euclidean distance $D(p, q)$ between voxels p and q . k is set to 3, which means 3 candidates with the smallest patch difference are selected in k -NN search. For other method (MV and LWV), we use their optimal parameters.

Fig. 3 plots the confidence map and five intermediate segmentation results of SC-LWV. In Fig. 3, we show the overall labeling confidence map, along with the segmentation results at the five stages that top 20%, 40%, 60%, 80% and all pixels (from column (b)~(f)) are labeled with the guidance of the confidence map C (column (a)). In this way, Fig. 3 validates that high-confidence pixels should guide the labeling procedure on low-confidence pixels, in order to increase the label accuracy, as proposed in our method.

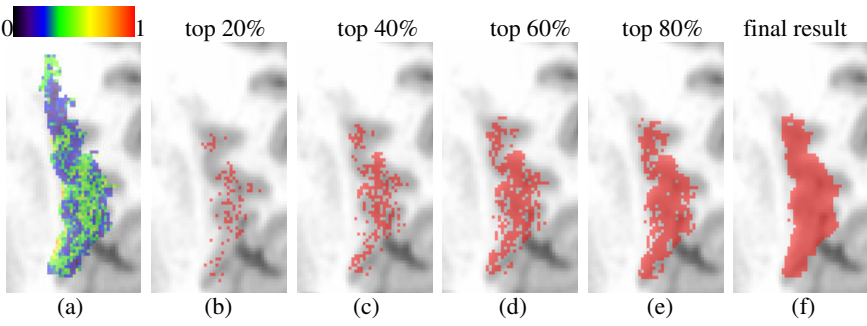


Fig. 3. Plots of the confidence map (a) and five intermediate segmentation results (b-f) using SC-LWV on *Left insula gyrus* segmentation

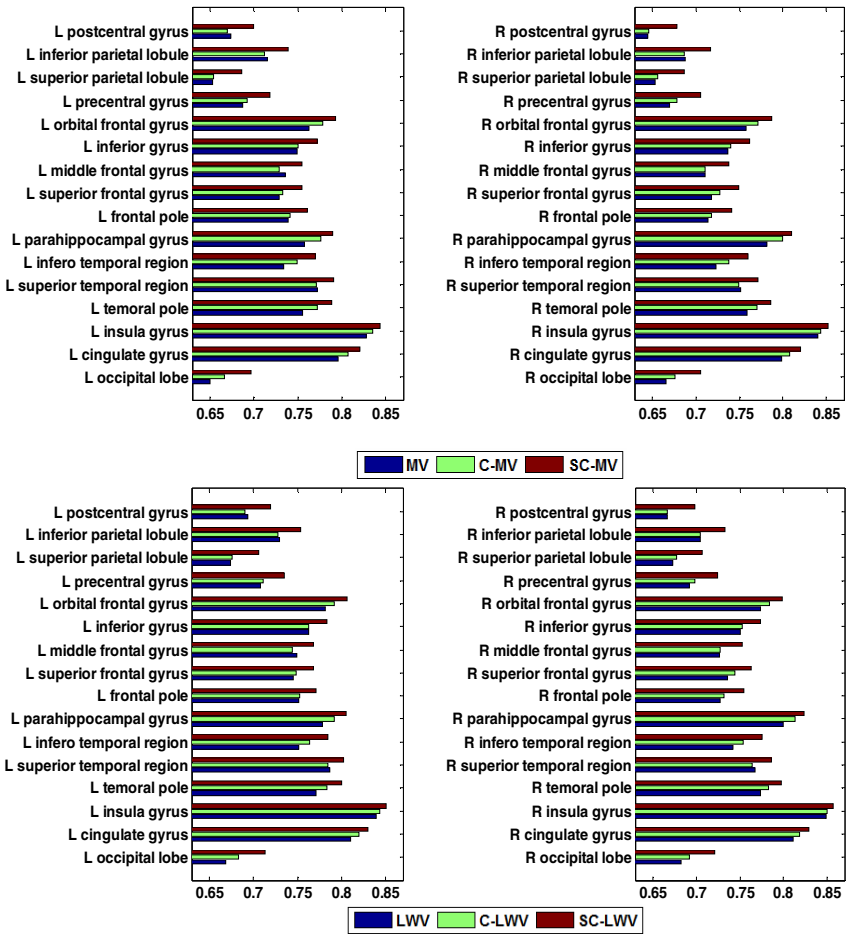


Fig. 4. Segmentation results on different brain structures measured by Dice overlap with six different label fusion algorithms

Fig. 4 shows the segmentation results on different brain structures (ROIs) using the six multi-atlas label fusion based segmentation algorithms. Here, the performance is measured with the Dice overlap, i.e., $Dice(S_a, S_b) = 2|S_a \cap S_b| / (|S_a| + |S_b|)$, where \cap indicates the overlapping voxels between the two segmentations, and $|S_a|$ indicates the number of voxels of the corresponding segmentation.

As we can see from Fig. 4, SC-MV and SC-LWV consistently improve the performance of MV and LWV, respectively, on all ROIs. For example, on ‘L occipital lobe’ ROI, SC-MV improves the overlap ratio from 0.650 (by MV) to 0.697, and SC-LWV improves the overlap ratio from 0.669 (by LWM) to 0.713. Fig. 4 also indicates that, in most cases, C-MV and C-LWV outperform MV and LWV respectively, but they are inferior to both SC-MV and SC-LWV in all cases. This demonstrates the importance of using the confidence-guided sequential labeling for ROI segmentation. Moreover, Fig. 4 shows that the LWV-based methods (LWV, C-LWV and SC-LWV) usually outperform the corresponding MV-based methods (MV, C-MV and SC-MV), which is consistent with previous studies [2, 4].

Finally, to have a summary on the segmentation accuracy on all 32 ROIs, we give the box plot for the results of the six algorithms, as shown in Fig. 5. It is observable that SC-MV and SC-LWV significantly improve the performance of the segmentation results by MV and LWV, respectively, while C-MV and C-LWV only slightly improve over the baseline methods (MV and LWV). Specifically, the averaged Dice overlap of MV, C-MV, SC-MV, LWV, C-LWV and SC-LWV are 0.730, 0.736, 0.758, 0.746, 0.751 and 0.772, respectively. Furthermore, the significant tests using paired t-test show that results of the proposed SC-MV and SC-LWV are significantly better than those of MV and LWV, respectively, at the 95% significance level. It validates the efficacy of the proposed sequential label fusion method.

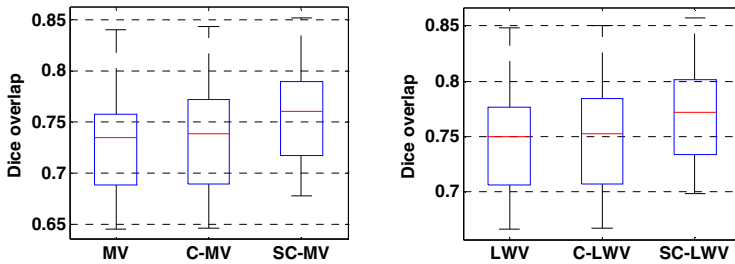


Fig. 5. Averaged Dice overlap using six different label fusion algorithms

4 Conclusion

We have presented a new label fusion method for multi-atlas based segmentation. In contrast to most existing label fusion methods which equally treat each voxel during label propagation, the proposed method considers the dependency among neighboring voxels for sequential labeling the voxels with confidence from high to low. To

achieve the sequential labeling, we define the labeling confidence which embeds not only the patch similarity but also the matching consistency to resolve the anatomy uncertainty in label fusion. Our method can be easily integrated with the current label fusion methods to significantly improve the label accuracy. Further work includes investigation of different estimations of label confidence, the effect of the sizes of local patch and neighborhood in label fusion, and more comparison with existing label fusion methods, e.g., STAPLE [10].

Acknowledgments. This work was supported in part by NIH grants EB006733, EB008374, EB009634 and MH088520, and also by National Science Foundation of China under grant No. 60875030.

References

1. Lotjonen, J., Wolz, R., Koikkalainen, J., Thurfjell, L., Waldemar, G., Soininen, H., Rueckert, D.: Fast and robust multi-atlas segmentation of brain magnetic resonance images. *Neuroimage* 49, 2352–2365 (2010)
2. Isgum, I., Staring, M., Rutten, A., Prokop, M., Viergever, M.A., van Ginneken, B.: Multi-atlas-based segmentation with local decision fusion-application to cardiac and aortic segmentation in CT scans. *IEEE Trans. Med.l Imag.* 28, 1000–1010 (2009)
3. Langerak, T.R., van der Heide, U.A., Kotte, A.N., Viergever, M.A., van Vulpen, M., Pluim, J.P.: Label fusion in atlas-based segmentation using a selective and iterative method for performance level estimation (SIMPLE). *IEEE Trans. Med.l Imag.* 29, 2000–2008 (2010)
4. Artachevarria, X., Munoz-Barrutia, A., de Solorzano, C.O.: Combination strategies in multi-atlas image segmentation: Application to brain MR data. *IEEE Trans. Med.l Imag.* 28, 1266–1277 (2009)
5. Aljabar, P., Heckemann, R.A., Hammers, A., Hajnal, J.V., Rueckert, D.: Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy. *Neuroimage* 46, 726–738 (2009)
6. Sabuncu, M.R., Yeo, B.T., Van Leemput, K., Fischl, B., Golland, P.: A generative model for image segmentation based on label fusion. *IEEE Trans. Med.l Imag.* 29, 1714–1729 (2010)
7. Coupe, P., Manjon, J.V., Fonov, V., Pruessner, J.: Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation. *Neuroimage* 54, 940–954 (2011)
8. Duda, R.O., Hart, P.E., Stork, D.G.: *Pattern Classification*. John Wiley and Sons, Inc., Chichester (2000)
9. Christensen, G., Geng, X., Kuhl, J., Bruss, J., Grabowski, T., Pirwani, I., Vannier, M., Allen, J., Damasio, H.: Introduction to the non-rigid Image registration evaluation project (NIREP). In: Pluim, J.P.W., Likar, B., Gerritsen, F.A. (eds.) *WBIR 2006*. LNCS, vol. 4057, pp. 128–135. Springer, Heidelberg (2006)
10. Warfield, S.K., Zou, K.H., Wells, W.M.: Simultaneous truth and performance level estimation (STAPLE): An algorithm for the validation of image segmentation. *IEEE Trans. Med.l Imag.* 23, 903–921 (2004)

Probabilistic Multi-shape Segmentation of Knee Extensor and Flexor Muscles

Shawn Andrews¹, Ghassan Hamarneh¹, Azadeh Yazdanpanah¹,
Bahareh HajGhanbari², and W. Darlene Reid²

¹ Medical Image Analysis Lab, Simon Fraser University, Canada
{sda56, hamarneh, aya18}@sfu.ca

² Department of Physical Therapy, University of British Columbia, Canada
wdreid@mail.ubc.ca, baharehg@interchange.ubc.ca

Abstract. Patients with chronic obstructive pulmonary disease (COPD) often exhibit skeletal muscle weakness in lower limbs. Analysis of the shapes and sizes of these muscles can lead to more effective therapy. Unfortunately, segmenting these muscles from one another is a challenging task due to a lack of image information in many areas. We present a fully automatic segmentation method that overcomes the inherent difficulties of this problem to accurately segment the different muscles. Our method enforces a multi-region shape prior on the segmentation to ensure feasibility and provides an energy minimizing probabilistic segmentation that indicates areas of uncertainty. Our experiments on 3D MRI datasets yield an average Dice similarity coefficient of 0.92 ± 0.03 with the ground truth.

1 Introduction

In patients with chronic obstructive pulmonary disease (COPD), skeletal muscle weakness is common [9, 11]. Lower limb muscles are often the most affected due to lack of use [11]. The skeletal muscle weakness related to COPD may be associated with the loss of muscle mass [11]. Reduced muscle mass is among several factors that lead to reduced force production, also including changes in the muscle contractile apparatus and neuromuscular activation. The relative effects of these factors are unknown [9]. Thus, size and shape measurements are required to study the contribution of muscle mass reduction to force loss. Recently [11], a non-uniform distribution of atrophy and size changes was found across knee extensors and flexors in patients with COPD, which may be reflective of localized factors such as denervation, limited recruitment, or atrophy of specific muscle fibers, rather than systemic factors contributing to muscle atrophy.

Magnetic resonance imaging (MRI) can be used to distinguish muscle from the surrounding regions, and can generate multiple image slices from which volume and shape properties of individual muscle can be estimated [6]. Information regarding the specific muscles or muscle regions that are most atrophied together with functional assessment will enable therapeutic interventions to be targeted to the affected regions [11, 14] rather than the prescription of a generalized approach that may prove ineffectual.

An important precursor to any volume measurement or shape analysis is segmentation. Manual segmentation in 3D medical images is extremely time consuming, tedious,

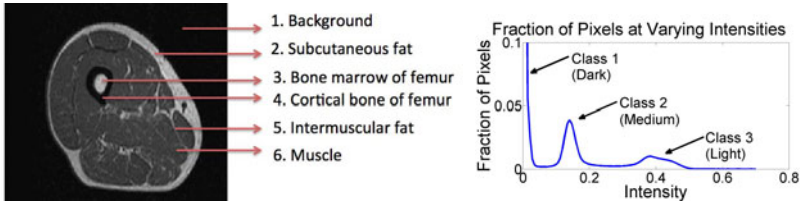


Fig. 1. Regions in the thigh, and the intensity profiles of the thigh MRIs. We see the background and cortical bone are dark, the muscle is medium, and the fat and bone marrow are light.

and suffers from inter- and intra-operator variability. Highly automated segmentation is important for studies involving a large cohort of subjects to reduce manual labour and variability, and to improve the efficiency in analyzing large groups of data.

Segmentation of thigh muscles is a challenging task, especially segmenting all 11 knee extensor and flexor muscles. As seen in Fig. 1, all these muscles have almost identical voxel intensities and weak inter-muscular image gradients. Similar muscle segmentation problems have been tackled previously using innovative techniques [15, 8, 7], but to the best of our knowledge there exists no work creating a fully automatic method for segmenting all 11 muscles from an MRI.

In this paper we present a segmentation method tailored to this specific problem. Given the difficulty of the problem, mainly due to the lack of image information in many regions, we generate a probabilistic segmentation in order to express confidence, or lack thereof, in these regions. In order to overcome the limited image information available, we will enforce a principal component analysis (PCA) based shape prior, constructed from training data, similar to [3]. Before PCA is performed on the probabilistic training segmentations, an anatomy-driven spatial alignment is performed and the isometric log-ratio (ILR) transform [2, 5] is applied to the training segmentations, thus moving them to a vector space to facilitate algebraic manipulation intrinsic to the probability simplex. A convex energy-minimization formulation is adopted that guarantees globally optimal solution insensitive to initialization.

2 Method

2.1 The ILR Transform

A probabilistic segmentation can be represented as a function $q : \Omega \rightarrow \mathbb{S}^R$, where Ω is the image domain and \mathbb{S}^R is the simplex of all valid $R = 12$ length probability vectors (11 muscles and background). If we define a function $\phi : \mathbb{S}^R \rightarrow \mathbb{R}^{R-1}$, bijectively mapping probability vectors to a vector of real numbers, then we can also represent probabilistic segmentations by another function $\eta = \phi \circ q : \Omega \rightarrow \mathbb{R}^{R-1}$. Using this representation, the space of segmentations, referred to as ILR space, is closed under linear combinations. This allows us to perform PCA on training segmentations without the need for constraints to ensure our results have valid probabilities, summing to unity.

We define ϕ as the ILR transform [2, 5]. In [5], the Aitchison inner product $\langle p, q \rangle_a$ and Aitchison distance metric $d_a(p, q)$ are defined for $p, q \in \mathbb{S}^R$. With the Aitchison

inner product, a basis for \mathbb{S}^R can be defined, denoted, $B_E = \{\mathbf{e}_1, \dots, \mathbf{e}_{R-1}\}$. There are many choices for B_E and some can be found in [5]; the exact choice of B_E does not affect our method. Given B_E , the ILR transform of $p \in \mathbb{S}^R$ is defined as:

$$\phi(p) = [\langle p, \mathbf{e}_1 \rangle_a, \dots, \langle p, \mathbf{e}_{R-1} \rangle_a] \in \mathbb{R}^{R-1}, \tag{1}$$

i.e. the projection of p onto the basis B_E . While there are several possible functions that map \mathbb{S}^R to a Euclidean space [13], we choose the ILR transform due to the fact that it is bijective (thus easily invertible) and an isometry. That is, for $p, q \in \mathbb{S}^R$, $d_a(p, q) = d(\phi(p), \phi(q))$, where $d(\cdot, \cdot)$ is the standard Euclidean distance. Finally, we discretize Ω to n voxels and represent segmentations η as a vectors of $n(R - 1)$ reals. We perform PCA on segmentations in this form to create our shape space.

2.2 The Shape Space

A strong shape prior that considers all of the muscles simultaneously is appropriate for this problem as the shape and relative locations of the muscles are highly correlated from one image to another. Furthermore, as the muscles are adjacent to each other along much of their boundaries, the shapes of different muscles in the same image are correlated. We construct such a shape prior by identifying a subspace of possible segmentations that corresponds to feasible thigh muscle segmentations and force our segmentation to lie in this space. To do this, we perform PCA on vectors $\{\hat{\eta}_1, \dots, \hat{\eta}_N\}$ in $\mathbb{R}^{n(R-1)}$, corresponding to training probabilistic ground truth (GT) segmentations.

We define η_0 as the mean of the training GTs, Γ as the $(n(R - 1)) \times k$ matrix whose columns correspond to the PCA eigenmodes of greatest variance, and Λ as the $k \times k$ diagonal matrix whose diagonal elements are the eigenvalues (variances) of the eigenmodes of Γ . We parameterize segmentations in our shape space by $\gamma \in \mathbb{R}^k$, so a segmentation is represented as

$$\eta(\gamma) = \eta_0 + \Gamma \gamma. \tag{2}$$

$\eta(\gamma)$ is linear in γ , so any energy functions that are convex in η will be convex in γ .

2.3 Image Alignment

In order to construct the shape space in Section 2.2, the thigh images must first be roughly aligned. To do this, we take advantage of some specific anatomical and image

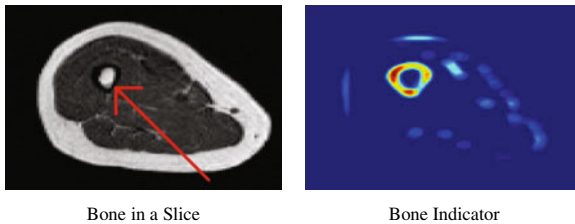


Fig. 2. A slice of a thigh image with the location of the bone highlighted (left, red arrow) and the locations of interfaces between medium and dark regions, indicating the boundary of the femur (right)

features. To bring the images into rough alignment, for each image we note that the images exhibit 3 distinctive intensity classes, referred to as “dark”, “medium”, and “light”. From Fig. 1 we see that background and cortical femur bone are dark, the muscles are medium, and the bone marrow and fat are light. We use these strong intensity priors to separate the muscles from the background and then crop the images accordingly.

To further align the images, we note the femur bone running through the middle of the thigh muscles follows a roughly straight vector through each image. As seen in Fig. 2 the bone is characterized by an interface between dark and medium intensity voxels, and such an interface occurs infrequently elsewhere in the image. For a given image, we estimate the vector along which the bone travels by finding its center, $B(z) = (x_B(z), y_B(z))$, in as many slices (z) as possible. We then fit a linear function $f(z) = (a_x z + b_x, a_y z + b_y)$ to estimate the coordinate of the centre of the bone in each slice. Using $f(z)$, we translate each slice in each image so that the femur follows the same vector in each image. Finally, we rescale the images isotropically so that their foregrounds are roughly the same size. This method is simple and quick, and we found it to provide an adequate, anatomy-driven alignment for construction of the shape space. A more complex alignment would unnecessarily shift work from our efficient convex segmentation method to an expensive non-convex registration scheme.

2.4 Energy Construction

With the images aligned and a shape space constructed, we will use image information to construct a strictly convex energy function over the shape space. The energy will be constructed such that viable segmentations yield lower energy values, and since it is strictly convex, the minimization of our energy will require no initialization.

As seen in Fig. 1 the different muscle regions have almost identical intensity distributions and textures, but the non-muscle region (fat and bone) has significantly different intensities. Thus, we create a probabilistic segmentation q_{BG} , where voxels with light or dark intensities are assigned to the non-muscle region with probability 1 and voxels with medium intensities are assigned probability $\frac{1}{11}$ for each of the 11 muscle regions. We then define an energy term:

$$E_{BG}(\gamma) = d(\eta_{BG}, \eta(\gamma))^2. \quad (3)$$

$E_{BG}(\gamma)$ is convex in η , and thus convex in γ , and encourages η to give high probability to the non-muscle region in appropriate places.

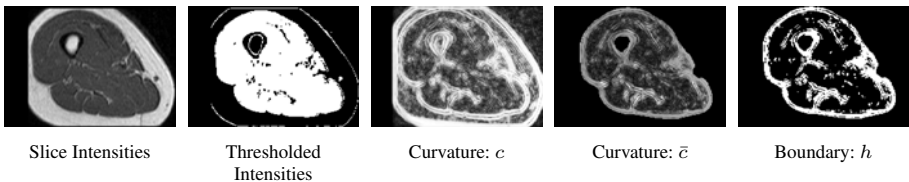


Fig. 3. Extracted image information. From left to right, we see for a thigh slice: the intensity values; the approximation of the muscle regions extracted by intensity thresholding; the curvature values c and \bar{c} ; and the boundary estimate h .

While the muscle regions are largely homogeneous in intensity, weak intensity gradients appear along the inter-muscular region boundaries (Fig. 3). The boundaries between muscle regions are characterized by a thin sheet of somewhat higher intensity fat voxels with lower intensity muscle voxels on either side, and therefore boundary voxels exhibit large curvature in the intensity in the direction perpendicular to the boundary. Thus we construct the Hessian of the intensity at each voxel x and find its eigenvalues, from largest to smallest magnitude, $c_1(x)$, $c_2(x)$, and $c_3(x)$, and the corresponding normalized eigenvectors $v_1(x)$, $v_2(x)$, and $v_3(x)$. $c_1(x)$ and $v_1(x)$ correspond to the maximum curvature and its direction. For voxels on these sheet boundaries, we expect $c_1(x) \gg c_2(x), c_3(x)$ [4]. Thus, we take $c(x) = \sqrt{c_1(x)^2 / (c_1(x)^2 + c_2(x)^2 + c_3(x)^2)}$ as a muscle boundary indicator at x (Fig. 3).

Once we have found c , we further filter out noise. To do this, we note that $v_1(x)$ for x on an intermuscular sheet of fat will be similar to $v_1(y)$ for a voxel y neighboring x within the sheet, and that the span of $v_2(x)$ and $v_3(x)$ will approximate the plane of the boundary. Thus we apply a specialized bilateral filter to $c(x)$, smoothing $c(x)$ with nearby voxels on the boundary, weighted by the dot product between their directions of maximal curvature. Specifically, if $y(\alpha, \beta) = x + \alpha v_2(x) + \beta v_3(x)$ then

$$\bar{c}(x) = \sum_{\alpha=-2}^2 \sum_{\beta=-2}^2 \frac{|\langle v_1(x), v_1(y(\alpha, \beta)) \rangle|}{\sqrt{1 + \alpha^2 + \beta^2}} c(y(\alpha, \beta)). \quad (4)$$

The boundary between the dark background and light subcutaneous fat does not help in the segmentation, so it is identified (based on intensity) and removed from \bar{c} (Fig. 3).

We construct \bar{c} for each training image and find c_{min} and c_{max} such that the set $H = \{x \mid c_{min} \leq \bar{c}(x) \leq c_{max}\}$ has the maximal Dice similarity coefficient (DSC) with the true set of boundary voxels (derived from the training GTs). Given a novel image, we construct \bar{c} , threshold to find H , and define $h : \Omega \rightarrow \{0, 1\}$ as the indicator function for H (Fig. 3). We use the boundary based energy term

$$E_{BDY}(\gamma) = \sum_{r=1}^{R-1} \sum_{x \in \Omega} (1 - h(x)) |\nabla_x \eta|^2. \quad (5)$$

E_{BDY} allows the segmentation to have large gradient only on voxels deemed likely to be part of a boundary, otherwise the gradient is penalized. Note once again that E_{BDY} is convex in η and thus in γ . Also note that $\nabla_x \eta$ measures the rate of change between the probabilities of neighboring voxels measured via the Aitchison distance.

Our final energy term enforces a Mahalanobis type penalty on γ , allowing components corresponding to eigenmodes of greater variance to vary more:

$$E_{Shape}(\gamma) = \gamma^T \Lambda^{-1} \gamma. \quad (6)$$

Our energy is a linear combination of the convex energy terms (3), (5), and (6):

$$E(\gamma) = \lambda_1 E_{BG}(\gamma) + \lambda_2 E_{BDY}(\gamma) + \lambda_3 E_{Shape}(\gamma). \quad (7)$$

Once our energy is constructed, we segment a novel image by finding $\gamma_{min} = \operatorname{argmin}_{\gamma} E(\gamma)$, and taking $q_{seg} = \phi^{-1}(\eta_0 + \Gamma \gamma_{min})$ as our final probabilistic segmentation.

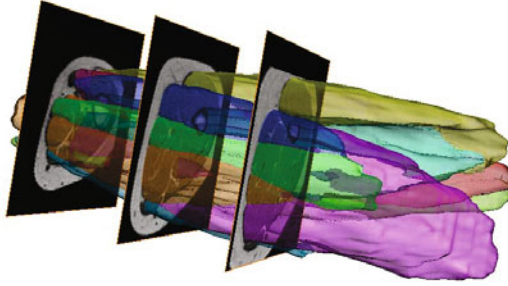


Fig. 4. (Color Figure) The resulting segmentation of our method overlaid on several image slices. Note the matching of the segmentation to the image boundaries.

3 Results

We report results on 3D MRI data from 40 subjects including 20 healthy subjects (11 women, age: 64.4 ± 8.1 years) and 20 moderate to severe COPD patients (11 women, age: 68 ± 10 years). A 1.5T MRI scanner acquired 5 mm-thick axial T1-weighted magnetic resonance (echo time, 8 ms; repetition time, 650 ms) with a 40 cm^2 field of view and a 512×384 pixel matrix (in-plane resolution, 0.78×1.78 mm). The MRI scan yielded a total of approximately 100 slices for each participant. To create the GT's, these images were manually segmented and each pixel was given a probability approaching 1 for the region it was assigned to.

We cropped the images down to a $175 \times 175 \times 85$ volume enclosing the thigh, and aligned the images using the method described in Section 2.3. We test our method using leave-one-out validation (removing one of the volumes and training a shape space using the remaining 39 volumes). Following [12], we increase the dimensions of the shape space by deforming each of the 39 training GT segmentations 3 times, giving 156 segmentations on which to train a shape space. Adding such deformed volumes to the training data will increase the expressibility of the shape space. We perform PCA on the training segmentations, finding the $k = 100$ eigenmodes of greatest variance.

We measured the success of our algorithm using the DSC between the segmentation we generate, thresholded to non-probabilistic, and the corresponding GT. In Fig. 4, we

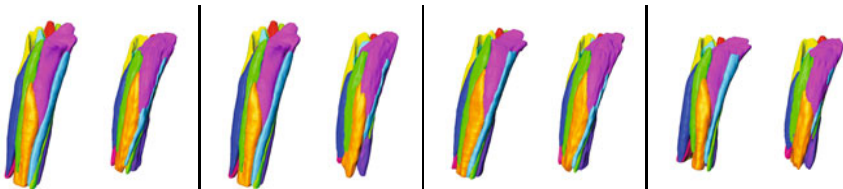


Fig. 5. (Color Figure) A comparison between several of the GT segmentations (left side of each pair) and the segmentations generated by our method (right side of each pair). The left two pairs are from COPD patients and the right two pairs are from non-COPD patients. DSC scores for each pair are 0.94, 0.91, 0.95, and 0.93.

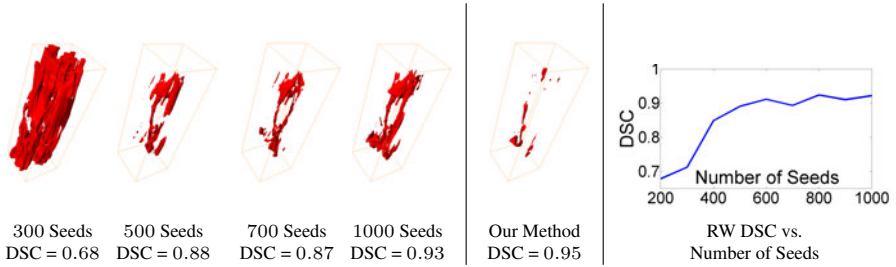


Fig. 6. The areas of mis-segmentation by the RW algorithm with varying numbers of randomly selected seeds (left) and our algorithm (center), along with a graph showing how the average DSC of RW changes with the number of seeds used (right)

see a segmentation generated by our method overlaid against slices of the image. We note our energy successfully matches the segmentation to boundaries in the image. We see examples of our segmentations compared to the GTs in Fig. 5 with DSC scores reported. We achieved an average DSC values of 0.92 ± 0.03 across all images.

To demonstrate the difficulty of the problem and the necessity of our shape prior, we segmented the thigh volumes using the popular random walker (RW) [10] algorithm, which also produces multi-region probabilistic segmentations and achieves the global minimum of its energy, but does not include a shape prior. RW requires seed voxels from each region to perform a segmentation, so we used the GT segmentations to provide between 200 and 1000 seeds, randomly sampled throughout the volume. In Fig. 6 we see the areas where the GT disagreed with our method and the various RW versions. RW only begins performing comparably to our method when the number of seed voxels exceeds 1000. In Fig. 6 we also see how the RW accuracy increases with the number of seeds provided. Note that without the GTs to provide automatic seeds, all seeds would have to be manually input, which would be very time consuming for a user. Also note that when 700 seeds were provided, the resulting DSC in Fig. 6 was worse than when only 500 seeds were provided, showing how the semi-automatic RW algorithm is dependent on seeding quality as well as quantity. In contrast, our method is strictly convex and fully automatic, not depending on user input.

4 Conclusion

The results of this work will be critical in clinical research aiming to determine factors affecting muscle structure and function in chronic conditions such as COPD. The methods proposed in this work will assist in more accurate measurement of muscle periphery and other size measures. Moreover it will provide researchers and clinicians with a time-efficient way to diagnose muscle anomalies associated with chronic conditions. For future work, we will explore projecting training GTs onto the shape space so that analysis of their eigenmode weights can be performed. These weights could both provide information about the adverse effects of COPD on the thigh muscle structure and help identify the disease in novel images. Our future work will also include exploring more advanced alignment techniques and comparing to other segmentation methods.

References

1. Bernard, S., LeBlanc, P., Whittom, F., Carrier, G., Jobin, J., Belleau, R., Maltais, F.: Peripheral muscle weakness in patients with chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.* 158(2), 629–634 (1998)
2. Changizi, N., Hamarneh, G.: Probabilistic multi-shape representation using an isometric log-ratio mapping. In: Jiang, T., Navab, N., Pluim, J., Viergever, M. (eds.) *MICCAI 2010. LNCS*, vol. 6363, pp. 563–570. Springer, Heidelberg (2010)
3. Cremers, D., Schmidt, F., Barthel, F.: Shape priors in variational image segmentation: Convexity, lipschitz continuity and globally optimal solutions. In: *CVPR 2008*, pp. 1–6. IEEE, Los Alamitos (2008)
4. Descoteaux, M., Audette, M., Chinzei, K., Siddiqi, K.: Bone enhancement filtering: Application to sinus bone segmentation and simulation of pituitary surgery. *Comput. Aided Surg.* 11(5), 247–255 (2006)
5. Egozcue, J., Pawlowsky-Glahn, V., Mateu-Figueras, G., Barceló-Vidal, C.: Isometric logratio transformations for compositional data analysis. *Mathematical Geology* 35, 279–300 (2003)
6. Engstrom, C., Loeb, G., Reid, J., Forrest, W., Avruch, L.: Morphometry of the human thigh muscles. A comparison between anatomical sections and computer tomographic and magnetic resonance images. *Journal of Anatomy* 176, 139–156 (1991)
7. Essafi, S., Langs, G., Deux, J.F., Rahmouni, A., Bassez, G., Paragios, N.: Wavelet-driven knowledge-based mri calf muscle segmentation. In: *Proceedings of the Sixth IEEE International Conference on Symposium on Biomedical Imaging: From Nano to Macro, ISBI 2009*, pp. 225–228. IEEE Press, Los Alamitos (2009)
8. Gilles, B., Magnenat-Thalmann, N.: Musculoskeletal mri segmentation using multiresolution simplex meshes with medial representations. *Medical Image Analysis* 14(3), 291–302 (2010)
9. Gosselink, R., Troosters, T., Decramer, M.: Peripheral muscle weakness contributes to exercise limitation in COPD. *Am. J. Respir. Crit. Care Med.* 153(3), 976–980 (1996)
10. Grady, L.: Multilabel Random Walker Image Segmentation Using Prior Models. In: *CVPR 2005*, pp. 763–770. IEEE, Los Alamitos (2005)
11. HajGhanbari, B., Hamarneh, G., Changizi, N., Ward, A., Reid, W.: MRI-Based 3D Shape Analysis of Thigh Muscles: Patients with Chronic Obstructive Pulmonary Disease Versus Healthy Adults. *Acad. Radiol.* 18(2), 155–166 (2011)
12. Hamarneh, G., Jassi, P., Tang, L.: Simulation of ground-truth validation data via physically- and statistically-based warps. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) *MICCAI 2008, Part I. LNCS*, vol. 5241, pp. 459–467. Springer, Heidelberg (2008)
13. Pohl, K.M., Fisher, J., Bouix, S., Shenton, M., McCarley, R.W., Grimson, W.E.L., Kikinis, R., Wells, W.M.: Using the logarithm of odds to define a vector space on probabilistic atlases. *Medical Image Analysis* 11(5), 465–477 (2007)
14. Tracy, B., Ivey, F.M., Metter, E.J., Fleg, J.L., Siegel, E.L., Hurley, B.F.: A more efficient magnetic resonance imaging-based strategy for measuring quadriceps muscle volume. *Med. Sci. Sports Exerc.* 35(3), 425–433 (2003)
15. Wang, C., Teboul, O., Michel, F., Essafi, S., Paragios, N.: 3d knowledge-based segmentation using pose-invariant higher-order graphs. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) *MICCAI 2010. LNCS*, vol. 6363, pp. 189–196. Springer, Heidelberg (2010)

Robust Skull Stripping of Clinical Glioblastoma Multiforme Data

William Speier, Juan E. Iglesias, Leila El-Kara, Zhuowen Tu, and Corey Arnold

University of California, Los Angeles, USA

Abstract. Skull stripping is the first step in many neuroimaging analyses and its success is critical to all subsequent processing. Methods exist to skull strip brain images without gross deformities, such as those affected by Alzheimer’s and Huntington’s disease. However, there are no techniques for extracting brains affected by diseases that significantly disturb normal anatomy. Glioblastoma multiforme (GBM) is such a disease, as afflicted individuals develop large tumors that often require surgical resection. In this paper, we extend the ROBEX skull stripping method to extract brains from GBM images. The proposed method uses a shape model trained on healthy brains to be relatively insensitive to lesions inside the brain. The brain boundary is then searched for potential resection cavities using adaptive thresholding and the Random Walker algorithm corrects for leakage into the ventricles. The results show significant improvement over three popular skull stripping algorithms (BET, BSE and HWA) in a dataset of 48 GBM cases.

1 Introduction

Automatic whole-brain extraction (known as skull stripping) from magnetic resonance images (MRI) is the first element of most neuroimaging pipelines. Therefore, its robustness is critical for the overall performance of the system. Many methods have been proposed to solve the problem and a good level of segmentation accuracy (overlap over 90% [1]) can usually be achieved with some parameter tuning. However, these methods are designed for MRI scans of brains that are either healthy or suffering from a disease that does not considerably alter the structure of the brain, such as Alzheimer’s disease, Huntington’s or multiple sclerosis. The segmentation accuracy drops substantially for scans of brains with tumors, contrast agents and/or resection cavities, which is typically the case in patients with glioblastoma multiforme (GBM).

The delineation of brains and resection cavities in volumes from patients with GBM is important in understanding the progression of the disease. Gross changes in morphology due to these processes have considerable effects on a patient’s quality of life and ultimate outcome. Skull stripping ensures that non-brain voxels do not erroneously contribute to any subsequent quantitative analysis, and also improves registration accuracy to a template or related studies.

Most current methods require a voxel intensity distribution close to that of a normal brain. For example, the widely-used Brain Surface Extractor (BSE) [2] uses diffusion filtering, edge detection and a chain of morphological operations

to segment the brain. This approach filters in images containing cancer-related findings which have strong borders that mislead the edge detection. Another approach is the Hybrid Watershed Algorithm (HWA) [3], part of the Freesurfer package, which combines a watershed algorithm, deformable surface and probabilistic atlas. This method has trouble with GBM images because the watershed algorithm assumes white matter connectivity and the atlas assumes a healthy brain shape. The Brain Extraction Tool (BET) [4] uses a deformable model grown based on voxel intensity. Cancer-related deformities hinder the evolution of the mesh and false positive regions appear around the eyes and brain stem. This problem may be ameliorated using a two-pass scheme (BET*) where the preliminary mask is used to align the brain to an atlas for a second pass by BET.

To the best of our knowledge, no existing technique robustly extracts the brain from MR images of individuals afflicted with GBM. The contribution of this paper is a method of segmenting brains with tumors, contrast agents and resection cavities to characterize GBM patients in support of research and clinical care. We address this problem by extending a robust skull stripping algorithm (ROBEX) [5] that extracts the brain using a strong prior on normal anatomy. We augment this method using convexity analysis, constrained distance transform and the Random Walker algorithm [6] to accurately find the brain boundary in the presence of tumors and surgical resection cavities. The proposed technique outperforms BSE, BET and HWA on a set of 48 volumes of subjects with GBM.

2 Materials: The Labeled MRI Dataset

In this work, 113 heterogeneous T1-weighted MRI volumes from individuals afflicted by GBM were collected. Under the supervision of a neuroradiologist, the brain boundary (including the cerebellum) was manually delineated by the third author in 48 of these cases. Of the 48 cases, 27 contained contrast enhancement, 23 contained resection cavities and 28 contained active tumor. The remaining 65 scans were only used for tuning parameters by visual inspection. The scans were acquired with several acquisition protocols with diverse voxel sizes. In-plane resolution (axial in all cases) ranges from 0.86 to 0.94mm. The slice separation was 3mm in 40 cases and 6mm in the other eight labelled cases.

Figure 1a shows a slice from a scan in our dataset that illustrates the difficulty of the extraction problem. At a glance, the resection cavity and left ventricle appear to be joined when the two are actually separated by the intact ventricular lining. Automatically delineating this resection-ventricle border is critical to understanding the subsequent morphological evolution after surgical resection.

3 Methods

The steps of the proposed method are illustrated in Figure 1. First, an initial brain boundary is obtained using ROBEX (Fig. 1b). Then, resection cavity candidates are found using adaptive thresholding (Fig. 1c). Each candidate is classified as either “cavity/undersegmentation” or “cavity and ventricle” using

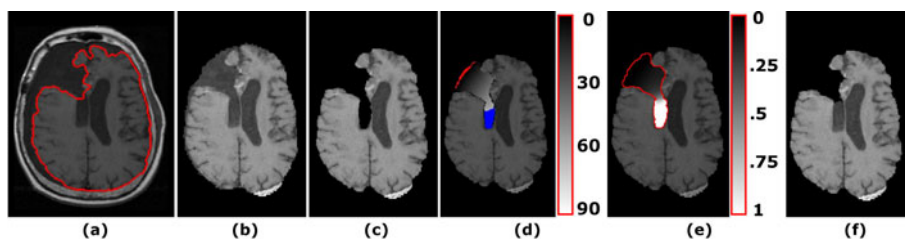


Fig. 1. Outline of the proposed algorithm. a) Axial slice of a T1 MRI with a resection cavity. b) Output from ROBEX, which leaves the cavity in the segmentation. c) Output from adaptive thresholding. d) Constrained distance transform within cavity/ventricle (in mm) and resulting seeds for Random Walker (positive in blue, negative in red). e) Probability map from Random Walker within the cavity/ventricle (marked in red). f) Final output, obtained by thresholding e).

convexity analysis. In the former case, the whole candidate region is removed from the brain mask. In the latter, a constrained distance transform (Fig. 1d) is used to seed Random Walker (Fig. 1e) which separates the region into cavity and ventricle (Fig. 1f). These steps are now discussed in detail.

3.1 Robust Brain Extraction

For the sake of completeness, a short description of ROBEX is included here. Designed to work with T1-weighted MRI, ROBEX applies signal intensity standardization and bias field correction to the input scan and then feeds it to a random forest classifier [7]. The random forest is trained to identify voxels on the brain surface. When applied, it generates a volume in which intensity represents the log-likelihood that a voxel is located on the boundary. Then, a shape model [8] of brain is fit to the output, maximizing the sum of the log-likelihood over the fit surface. This output is refined by allowing a small free deformation outside the model, which is optimized using the max-flow min-cut algorithm [9]. A sample output from ROBEX is displayed in Figure 1b.

ROBEX is very robust against tumors in the brain because it focuses on finding the brain boundary. However, contrast agents used in brain cancer imaging can bias the intensity standardization step as it is based on min/max normalization. We solve this problem by cropping the rightmost 5% of the histogram when computing the maximum. This value is high enough to remove all hyper-intense tumor enhancement and residual skull, but low enough to retain some of the white matter. With regard to resection cavities, ROBEX will close the mask around them in most cases (unless they are very shallow) due to the fact that it is heavily based on of model of healthy brain. A solution to this problem is described in the following sections.

3.2 Adaptive Thresholding

Resection cavities correspond to low-intensity voxels in T1-weighted MRI because they are filled with cerebrospinal fluid (CSF). A simple thresholding

operator typically suffices to separate most of the CSF voxels from the rest of the brain matter, which is the principle behind some skull stripping methods [10,11]. However, algorithms of this type may fail in the presence of tumors and contrast agents that can adversely influence the global threshold. Instead, we use an adaptive scheme in which τ , the threshold value, is space-dependent:

$$\tau(x, y, z) = \frac{\sigma_{glob}^2}{\sigma_{loc}^2(x, y, z) + \sigma_{glob}^2} \tau_{glob} + \frac{\sigma_{loc}^2(x, y, z)}{\sigma_{loc}^2(x, y, z) + \sigma_{glob}^2} \tau_{loc}(x, y, z)$$

where σ_{loc}^2 is the local image intensity variance computed in a sphere of radius r_{loc} around each pixel, σ_{glob}^2 is the global variance across the scan within the preliminary brain mask, and τ_{glob} and τ_{loc} are a global and a local thresholds. The global threshold is the average of the median intensity in the preliminary brain mask and the minimum after cropping the leftmost 5% of the histogram. The local threshold is the mean intensity in the sphere around each pixel. Compared to using τ_{glob} alone, incorporating a local threshold produces a more accurate boundary, especially around resection cavities. After thresholding, the mask is smoothed by morphological closing with a spherical element (radius 5 mm), filling holes and removing islands. Sample output is displayed in Figure 11c.

3.3 Detecting Cavities

The next step in the pipeline is detecting resection cavities left in the segmentation by ROBEX. For each connected component of the negated mask, a shape convexity index $c \in (0, 1]$ is computed as the ratio between its volume and that of its convex hull. The components for which $c > c_{min}$ are assumed to correspond to resection cavities, and the output mask is set to zero in that region. If $c < c_{min}$, leakage into the ventricles is suspected, and the candidate is fed to the next step of the pipeline. It is fairly easy to find a value of c_{min} that separates the two classes because the convexity drops considerably in the ventricles.

3.4 Separating Resection Cavities from Ventricles

Although a resection cavity and the ventricles are not connected in reality, it is often difficult to discern the border between the two due to the lack of contrast (both contain CSF), noise and limitations in resolution. Given a set of pre-labeled voxels (seeds), Random Walker can make this delineation by determining the probability of randomly reaching a seed from each unlabeled voxel.

Random Walker was conceived as a semi-automated segmentation method with manually placed seeds, but this work uses automatic seeding. A natural choice for the negative seed is the intersection of the cavity/ventricle candidate and the boundary of the mask of ROBEX, which is the most likely region to be part of the cavity. For the positive seed, we selected all points located more than d_{min} away from the negative seeds using a constrained distance transform [12]. The seed regions are overlaid in Figure 11d, the probability of resection cavity in the candidate region is shown in Figure 11e and the final segmentation (obtained by thresholding that probability at 0.5) is displayed in Figure 11f.

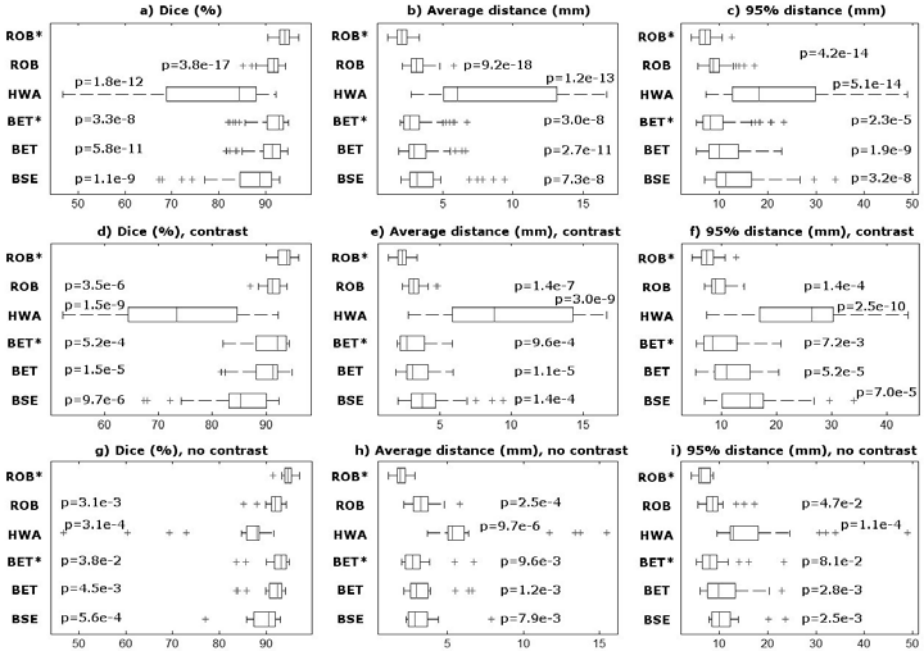


Fig. 2. Box plots and p-values for the different methods. On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles and the whiskers extend to the most extreme data points not considered outliers i.e. within three standard deviations from the mean. The outliers are plotted individually. The p-values correspond to paired one tailed t-tests between our method and all of the others. ROB refers to ROBEX and ROB* to the extension proposed here.

4 Experiments and Results

4.1 Experimental Setup

As described in section 2, the 65 unlabeled scans were first used in pilot experiments to design the algorithm and tune parameters by visual inspection of the output. $r_{loc} = 4\text{mm}$, $c_{min} = 0.7$ and $d_{min} = 60\text{mm}$ were found to be good values. The 48 scans for which ground truth was available were skull-stripped using BET v2.1 (one and two passes), BSE v09, HWA from FreeSurfer 5.0.0, ROBEX and the proposed extension. Default parameter values were used for BET, BSE and HWA. Three different metrics were used to measure the performance of the different methods: 1) the Dice overlap index $DICE(X, Y) = \frac{2|X \cap Y|}{|X| + |Y|}$; 2) the mean symmetric surface-to-surface distance; and 3) the 95% percentile of the symmetric surface-to-surface distance, a robust alternative to the Hausdorff (i.e., maximal) distance. Measure (3) is an effective complement to (1) and (2) as it captures fine boundary detail as opposed to gross overlap.

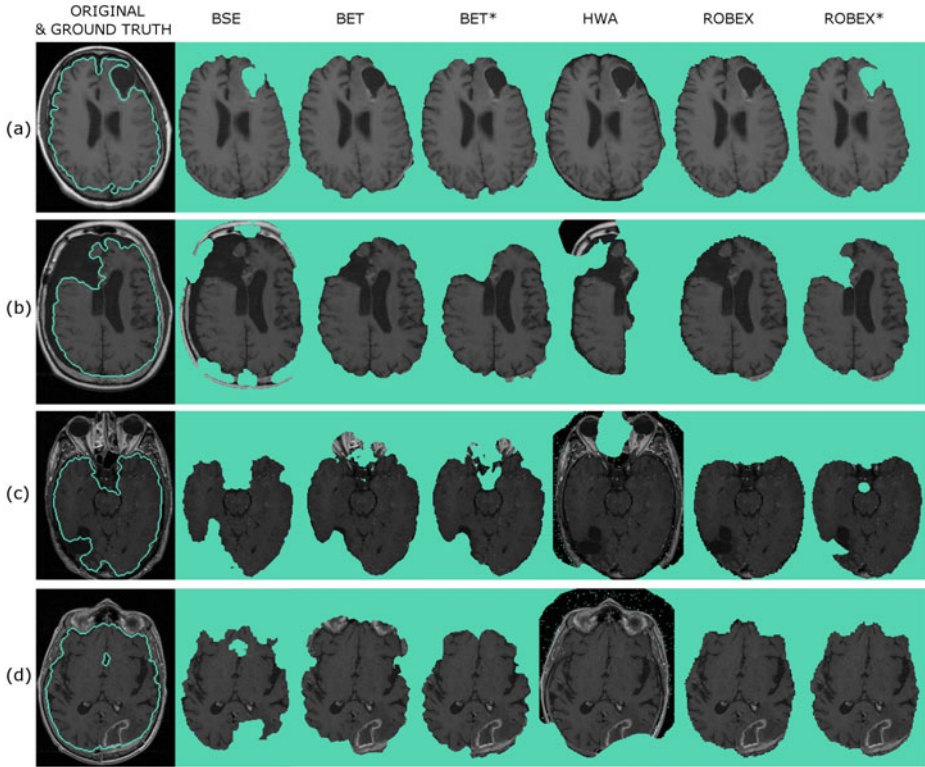


Fig. 3. Axial slices from sample outputs for the different extraction methods. (a) Subject with resection cavity. (b-c) Subjects with resection cavities that interface with ventricles. (d) Subject with an unresected tumor (note the hyperintense region around the tumor due to contrast). ROBEX* represents the extension of ROBEX.

4.2 Results

Fig. 2 displays the box plots for each evaluation metric for all methods as well as the p-values corresponding to one tailed t-tests comparing the proposed method to the other approaches. When all test scans are considered (Fig. 2a-c), our method significantly outperforms the selected comparative methods in terms of the three metrics ($p \in [9.2e - 18, 2.3e - 5]$). Moreover, the robustness of our method is apparent from the box plots. For example, the minimum DICE overlap across the dataset is 67.3% for BSE, 81.6% for BET, 82.1% for two-pass BET, 46.7% for HWA, 85.2% for ROBEX and 90.1% for the proposed algorithm. When the scans with (Fig. 2d-f) and without contrast (Fig. 2g-i) are considered separately, we observe that the improvement with respect to the other methods is larger when contrast agents are present. However, in cases with no contrast, the difference is still significant at $p = 0.05$ for all the methods and metrics except for the 95% percentile of the distance of 2-pass BET.

Fig. 3 displays the output of the different methods for a few examples. BSE can reasonably handle resection cavities in some cases (Fig. 3a), but often fails to completely extract the brain boundary (Fig. 3b). BET consistently delineates the brain surface (Fig. 3b), but has difficulty identifying resection cavities as the deformable surface can: 1) grow beyond the cavity (Fig. 3a) or; 2) have insufficient flexibility to surround it (Fig. 3c). In addition, its known limitation of undersegmentation around the eyes is apparent (Fig. 3c-d). Two-pass BET improves the results from BET and has the second-highest performance after our method. HWA does not adapt well to the dataset as its watershed algorithm relies heavily on white matter intensity estimation, which is disturbed by tumors, contrast enhancement and resection cavities. Finally, the proposed method succeeds at separating resection cavities from brain while not removing unresected tumors. In addition, it often produces a good approximation of the cavity-ventricle interface (Fig. 3b-c), which the other methods seldom do.

5 Discussion and Conclusion

A skull stripping method that can handle scans from subjects with an aggressive cancer (GBM) has been presented in this paper. Experimental evaluation on 48 subjects shows that our method outperforms three other widely-used brain extraction algorithms: BET, BSE and HWA. While we acknowledge that these techniques were not designed for this specific problem, to the best of our knowledge they represent the best alternative options. Since the proposed post-processing requires a brain-like initialization, it is applicable only to ROBEX.

The proposed method is specific to T1-weighted MRI. When several channels are acquired (e.g. T1, T2, proton density, pre- and post-contrast), skull stripping is typically performed on the T1 volume and the resulting mask is propagated to the other channels. In fact, most brain extraction methods in the literature (BET being a notable exception) are designed to work with T1 data. However, it would certainly be possible to take advantage of multiple modalities to improve the stripping results. This direction remains to be explored.

An interesting by-product of the proposed algorithm is the segmentation of resection cavities (see Fig. 1e). Finding non-ventricular CSF is a common processing step after skull stripping in neuro image analysis pipelines, but existing methods do not handle the case where CSF might be in resection cavities adjacent to the ventricles. It would be interesting to evaluate to what extent the segmentation given by our method is reliable, since it may be useful to quantify the evolution of the resection. Exploring this direction and evaluating the algorithm on more cases also remain as future work.

Finally, it is important to discuss the algorithms' computational complexity. BET and BSE run in a few seconds, whereas ROBEX and HWA require one or two minutes on a modern machine. The proposed post-processing only requires a few seconds, which is almost negligible next to the execution time of ROBEX.

Acknowledgements. This work was funded by NSF grant 0844566 and NIH grants R01-LM009961 and T15-LM007356.

References

1. Fennema-Notestine, C., Ozyurt, I., Clark, C., Morris, S., Bischoff-Grethe, A., Bondi, M., Jernigan, T., Fischl, B., Segonne, F., Shattuck, D., Leahy, R., Rex, D., Toga, A., Zou, K., BIRN, Brown, G.: Quantitative evaluation of automated skull-stripping methods applied to contemporary and legacy images: effects of diagnosis, bias correction, and slice location. *Human Brain Mapping* 27(2), 99–113 (2006)
2. Shattuck, D., Sandor-Leahy, S., Schaper, K., Rottenberg, D., Leahy, R.: Magnetic resonance image tissue classification using a partial volume model. *NeuroImage* 13(5), 856–876 (2001)
3. Segonne, F., Dale, A., Busa, E., Glessner, M., Salat, D., Hahn, H., Fischl, B.: A hybrid approach to the skull stripping problem in MRI. *Neuroimage* 22(3), 1060–1075 (2004)
4. Smith, S.: Fast robust automated brain extraction. *Human Brain Mapping* 17(3), 143–155 (2002)
5. Iglesias, J., Liu, C., Thompson, P., Tu, Z.: Robust brain extraction across datasets and comparison with publicly available methods. *IEEE Transactions on Medical Imaging* (in press, 2011)
6. Grady, L.: Random walks for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28(11), 1768–1783 (2006)
7. Breiman, L.: Random forests. *Machine Learning* 45(1), 5–32 (2001)
8. Cootes, T., Taylor, C., Cooper, D., Graham, J.: Active shape models-their training and application. *Computer Vision and Image Understanding* 61(1), 38–59 (1995)
9. Li, K., Wu, X., Chen, D., Sonka, M.: Optimal surface segmentation in volumetric images—a graph-theoretic approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28(1), 119–134 (2006)
10. Sadanathan, S., Zheng, W., Chee, M., Zagorodnov, V.: Skull stripping using graph cuts. *NeuroImage* 49(1), 225–239 (2010)
11. Mikheev, A., Nevsky, G., Govindan, S., Grossman, R., Rusinek, H.: Fully automatic segmentation of the brain from T1-weighted MRI using Bridge Burner algorithm. *Journal of Magnetic Resonance Imaging* 27(6), 1235–1241 (2008)
12. Piper, J., Granum, E.: Computing distance transformations in convex and non-convex domains. *Pattern Recognition* 20(6), 599–615 (1987)

Multi-stage Learning for Robust Lung Segmentation in Challenging CT Volumes

Michal Sofka¹, Jens Wetzl¹, Neil Birkbeck¹, Jingdan Zhang¹,
Timo Kohlberger¹, Jens Kaftan², Jérôme Declerck², and S. Kevin Zhou¹

¹ Image Analytics and Informatics, Siemens Corporate Research, Princeton, NJ, USA

² Molecular Imaging, Siemens Healthcare, Oxford, UK

Abstract. Simple algorithms for segmenting healthy lung parenchyma in CT are unable to deal with high density tissue common in pulmonary diseases. To overcome this problem, we propose a multi-stage learning-based approach that combines anatomical information to predict an initialization of a statistical shape model of the lungs. The initialization first detects the carina of the trachea, and uses this to detect a set of automatically selected stable landmarks on regions near the lung (e.g., ribs, spine). These landmarks are used to align the shape model, which is then refined through boundary detection to obtain fine-grained segmentation. Robustness is obtained through hierarchical use of discriminative classifiers that are trained on a range of manually annotated data of diseased and healthy lungs. We demonstrate fast detection (35s per volume on average) and segmentation of 2 mm accuracy on challenging data.

1 Introduction

Lung segmentation in thoracic CT images is an important prerequisite for detection and study of the progression and treatment of pulmonary diseases. Due to their high air content, healthy lung has lower attenuation than the surrounding tissue, allowing easy detection through standard thresholding and region-growing methods (e.g., [2]). However, pulmonary diseases (e.g., pulmonary fibrosis) lead to higher density tissue, and cause a changed appearance (e.g., different texture), making it hard to segment robustly (Figure 1).

In this paper, we present an effective learning-based segmentation technique that addresses the changes in lung appearance due to pulmonary diseases. The first step of the algorithm is the robust detection of the carina of the trachea with a discriminative classifier. The carina location is used to *predict* approximate poses (translation, orientation, and size) of the left and right lung. The prediction is based on a prior model obtained from a large expert-annotated database of

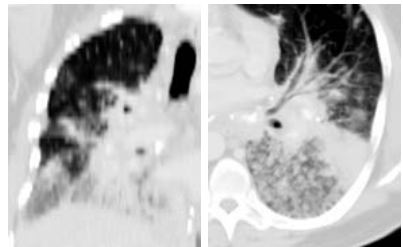


Fig. 1. Pulmonary diseases lead to higher density tissue which complicates standard segmentation algorithms

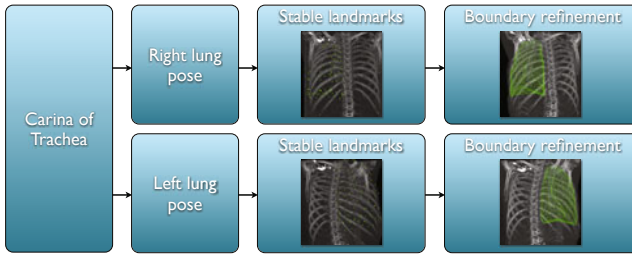


Fig. 2. System diagram: Carina detection allows the prediction of lung poses, which give initial locations for stable landmarks on the lung surface. That surface is then refined with a boundary detector.

lung scans. Placing a mean lung shape into the bounding box implied by each pose gives initial locations of a set of stable landmarks, which are selected automatically during training using the uncertainty of their locations. The locations of the landmarks are then locally refined with robust detectors. The final refinement is performed by a boundary detector which accurately estimates the lung surface. The overall system diagram is shown in Figure 2.

By focusing on stable landmarks and progressing from a coarse set to a fine set, we rely on the local regions which are most consistent, even in the presence of abnormalities. Stable landmarks are typically selected near vertebrae, ribs, and other distinctive anatomical structures (Fig. 4). In detection, the stability of the landmarks is further guaranteed by using a discriminative classifier (PBT [10]) which includes a powerful feature selection operating on a feature pool computed from all training volumes. Features susceptible to abnormal diseased areas are therefore not selected for landmark detection.

Existing approaches to increase robustness of lung segmentation focus on specific pathologies [11, 4], rely on interaction [3], adapt the simple thresholding to regions that often complicate lung segmentation [1], or augment texture cues for interstitial lung disease [12]. Such methods are not capable of handling the moderate to extreme pathologies that exhibit higher density tissue. More elaborate methods use anatomical information [6], shape priors [9], or statistical methods to detect patterns in the diseased tissue [7]. Shape models alone with simple image cues [9] are not enough to provide robustness to change in tissue density, and the anatomical constraints, priors or machine learning techniques need to be combined. To date, few methods combine either shape constraints or anatomical information with learning [7] for lung segmentation, but do so with simple classifiers (e.g., k-nearest neighbor) and use limited features.

In this work, we combine a statistical model of shape variation with statistical pattern recognition that uses anatomical information for robust lung segmentation. A wide gamut of gradient and intensity features capable of discriminating diseased lung tissue and implicitly capable of encoding anatomical relationships is selected by a powerful discriminative classifier, the probabilistic boosting tree (PBT) [10]. The classifiers and shape model are trained on a database of normal

and diseased tissues. Through fast coarse detection and refinement based on a hierarchical detection network (HDN) [8], a segmentation is obtained in 35 seconds. We demonstrate our implementation on a number of challenging pathological thoracic CT images. The average error on unseen data is 1.98 mm for the right and 1.92 mm for the left lung.

2 Learning

The algorithm starts by detecting the carina of the trachea. Its location is then used in a Hierarchical Detection Network (HDN) [8] to predict pose parameters of left and right lung and subsequently initialize the set of stable landmarks. The landmark locations are then refined and used to provide a rough boundary estimate. This estimate is improved during boundary detection which results in the final accurate lung segmentation.

2.1 Hierarchical Detection Network

The hierarchical detection network estimates unknown object states (e.g., object poses) as a sequential decision process. The formulation is similar to Markov chain approaches to object tracking, but instead of a temporal motion model with temporal observations, there is a spatial dependence (or prior relationship) between objects. The unknown parameters of each object are denoted as θ_t (e.g., the 9 parameters of a similarity transform), and the complete state for $t + 1$ objects is denoted $\theta_{0:t}$. Estimation of each object utilizes an observation region of the input volume, $V_t \subseteq V$, where $V : \mathbb{R}^d \mapsto \mathbb{R}$ is d dimensional input. The posterior density of the complete state, $f(\theta_{0:t}|V_{0:t})$, is approximated through a sequence of recursive *prediction* and *update* steps.

The *prediction* approximates the detection up to object t using the transition probability, $f(\theta_t|\theta_{0:t-1})$, and the posterior up to object $t - 1$:

$$f(\theta_{0:t}|V_{0:t-1}) = f(\theta_t|\theta_{0:t-1})f(\theta_{0:t-1}|V_{0:t-1}) \quad (1)$$

The *update* then fuses the results with the new observation region, V_t :

$$f(\theta_{0:t}|V_{0:t}) = \frac{f(V_t|\theta_t)f(\theta_{0:t}|V_{0:t-1})}{f(V_t|V_{0:t-1})} \quad (2)$$

The likelihood, $f(V_t|\theta_t)$, is empirically modeled by training a discriminative model. Concretely, letting $y \in \{-1, 1\}$ be a random variable denoting the occurrence of an object at pose θ_t , the likelihood is defined as:

$$f(V_t|\theta_t) = f(y = +1|V_t, \theta_t) \quad (3)$$

Where the posterior, $f(y = +1|V_t, \theta_t)$, is the output of a discriminative classifier (e.g., the probabilistic boosting tree [10]).

The transition prior approximates the sequential dependence of object t as a Gaussian distribution from one of the previous objects (Figure 2):

$$f(\theta_t|\theta_{0:t-1}) = f(\theta_t|\theta_j), \quad \exists j \in \{0, t-1\} \quad (4)$$

2.2 Pose Detection

In the case of pose detection in 3D, the state of each object is compactly represented with 9 parameters including the position $\mathbf{p} \in \mathbb{R}^3$, orientation as Euler angles \mathbf{r} , and scale, \mathbf{s} , of the object: $\boldsymbol{\theta}_t = \{\mathbf{p}_t, \mathbf{r}_t, \mathbf{s}_t\}$. For efficiency these three sets of parameters are treated as a chain of dependent estimates [5]:

$$f(\boldsymbol{\theta}_t|V_t) = f(\mathbf{p}_t|V_t)f(\mathbf{r}_t|\mathbf{p}_t, V_t)f(\mathbf{s}_t|\mathbf{p}_t, \mathbf{r}_t, V_t) \quad (5)$$

Splitting up the pose estimation in this way reduces the dimensionality of each sub-problem allowing fewer particles to be used during estimation.

2.3 Selection of Stable Landmarks

The set of stable landmarks is selected during training as follows. First, the annotation meshes are aligned to a common coordinate frame (see the next section). The correspondences formed during alignment identify each mesh vertex across all meshes (and volumes). The mesh vertices are used as landmark candidates. Denoting their location as $\{\mathbf{g}_i\}$. One position detector per landmark candidate is trained using all annotations. The detectors are then used to obtain detection results for each landmark, denoted as $\{\mathbf{d}_i\}$. The uncertainty of each detector is modeled by the covariance matrix, \mathbf{C}_i , of the final detected candidate location: $\mathbf{C}_i = \sum_i \mathbf{e}_i \mathbf{e}_i^\top$, where $\mathbf{e}_i = \mathbf{d}_i - \mathbf{g}_i$. The stable landmarks are selected one by one according to the score criterion $s_i = \text{trace}(\mathbf{C}_i)$ (higher s indicates higher uncertainty). During this selection, we apply spatial filtering (with radius $r = 20\text{mm}$) using the score s_i . This way, we obtain a set of stable landmarks with low uncertainty that are widely distributed along the lung surface.

2.4 Shape Initialization

After the poses of left and right lung have been detected, the boundary of the lung is detected to find an initial segmentation. This initial segmentation is a deformation of a triangulated mesh model. The model, $\mathcal{M} = (\mathcal{P}, \mathcal{T})$ consists of a set of points, $\mathcal{P} = \{\mathbf{x}_i \in \mathbb{R}^3\}_{i=1}^N$, and a set of triangle indices, $\mathcal{T} = \{\Delta_j \in \mathbb{Z}^3\}_{j=1}^M$.

The high dimensional search space is restricted by a prior learned linear model of shape variation. The prior shape model, $\mathcal{S} = (\{\hat{\mathbf{x}}\}_{i=1}^N, \{U_j\}_{j=1}^M)$, consists of a mean shape and a set of linear basis shapes, $U_j = \{u_{ij}\}_{i=1}^N$, that are learned through procrustes analysis and PCA on training data. A synthesized shape in the *span* of the shape-space can be specified by a few PCA coefficients, $\{\lambda_j\}$, and a pose, $(\mathbf{p}, \mathbf{r}, \mathbf{s})$:

$$g(\mathbf{x}_i; \{\lambda_j\}, \mathbf{p}, \mathbf{r}, \mathbf{s}) = \mathbf{p} + \mathbf{M}(\mathbf{s}, \mathbf{r}) \sum_j (\hat{\mathbf{x}}_i + \mathbf{u}_{ij} \lambda_j) \quad (6)$$

where $\mathbf{M}(\mathbf{s}, \mathbf{r})$ is a 3×3 scale and rotation matrix.

Estimation of the first three coefficients is done in the HDN framework, where $\boldsymbol{\theta}_t = \{\lambda_1, \lambda_2, \lambda_3\}$. Particles from the pose estimation process are augmented

with three PCA coefficients sampled uniformly over the range of coefficients observed in the training data. Similar to Eq. 3, the observation model, $f(\theta_t|V_t)$, is empirically modeled with a discriminative classifier that uses steerable features evaluated on surface points of the synthesized mesh 5.

2.5 Freeform Refinement

The first three PCA coefficients give a coarse approximation to the boundary. In order for the shape model to be expressive enough for all real instances, a large number of basis functions may be needed (e.g., the order of 100s). Instead of estimating all of the λ coefficients directly as above, the freeform refinement takes an iterative surface deformation approach 5.

Starting with the initialized shape from above, the freeform refinement seeks to find the most probable mesh, \mathcal{M} , in the space of the linear shape model:

$$\max f(\mathcal{M}|V_t) \quad s.t. \quad \mathcal{M} \in span(\mathcal{S}) \quad (7)$$

Where $f(\mathcal{M}|V_t)$ is approximated by integrating over the surface:

$$f(\mathcal{M}|V_t) = \frac{1}{N} \sum_{x_i} f(\mathbf{x}_i|V_t). \quad (8)$$

Here the per-point posterior is directly approximated by a discriminative model. Letting $y_i = \{-1, +1\}$ be a random variable denoting the presence of a surface at point \mathbf{x}_i along normal \mathbf{n}_i :

$$f(\mathbf{x}_i|V_t) = f(y_i = 1|\mathbf{x}_i, \mathbf{n}_i, V_t) \quad (9)$$

Instead of performing a coupled high dimensional optimization for all points simultaneously, local search within a predefined range $\{-\tau, \tau\}$ is performed for each vertex to find the best displacement along the normal, $\mathbf{x}_i \leftarrow \mathbf{x}_i + d_i \mathbf{n}_i$:

$$d_i = \arg \max_{-\tau \leq d \leq \tau} f(\mathbf{x}_i + d \mathbf{n}_i|V_t) \quad (10)$$

The resulting shape is projected onto the shape-space and surface normals are updated. This interleaved displace and regularization process is iterated several times. In latter iterations, τ is reduced, and the shape is allowed to vary from the $span(\mathcal{S})$. In these iterations, instead of regularizing by projecting into the shape space, a simple mesh smoothing is used to regularize the displaced mesh.

Table 1. Results of symmetrical point-to-mesh comparisons of detected results and annotations for both lungs, with and without stable landmark detection

Lung	Landmark	Mean (std.)	Med.	Min	Max	80%
right	no	2.35 ± 0.86	2.16	1.40	6.43	2.57
right	yes	1.98 ± 0.62	1.82	1.37	4.87	2.18
left	no	2.31 ± 2.42	1.96	1.28	21.11	2.22
left	yes	1.92 ± 0.73	1.80	1.19	6.54	2.15

3 Experiments

Our experiments start by analyzing the error of detectors during training. We then show the set of top automatically selected landmarks. Finally, we provide qualitative and quantitative evaluation of lung segmentation.

Our dataset consists of 260 expert-annotated diagnostic CT scans of varying contrast. The slice thickness varies from 0.5 to 5.0. The dataset is randomly separated into two disjoint sets, one for training (192 volumes) and one for testing (68 volumes).

The first result in Figure 3 shows the sorted errors of all candidate landmarks (Section 2.3). These are the detection results obtained from landmark detectors

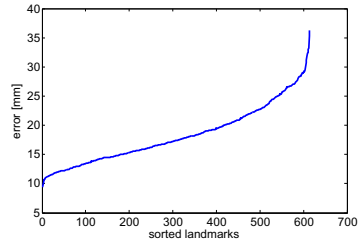


Fig. 3. Sorted mean errors of the 614 landmarks computed from all training volumes

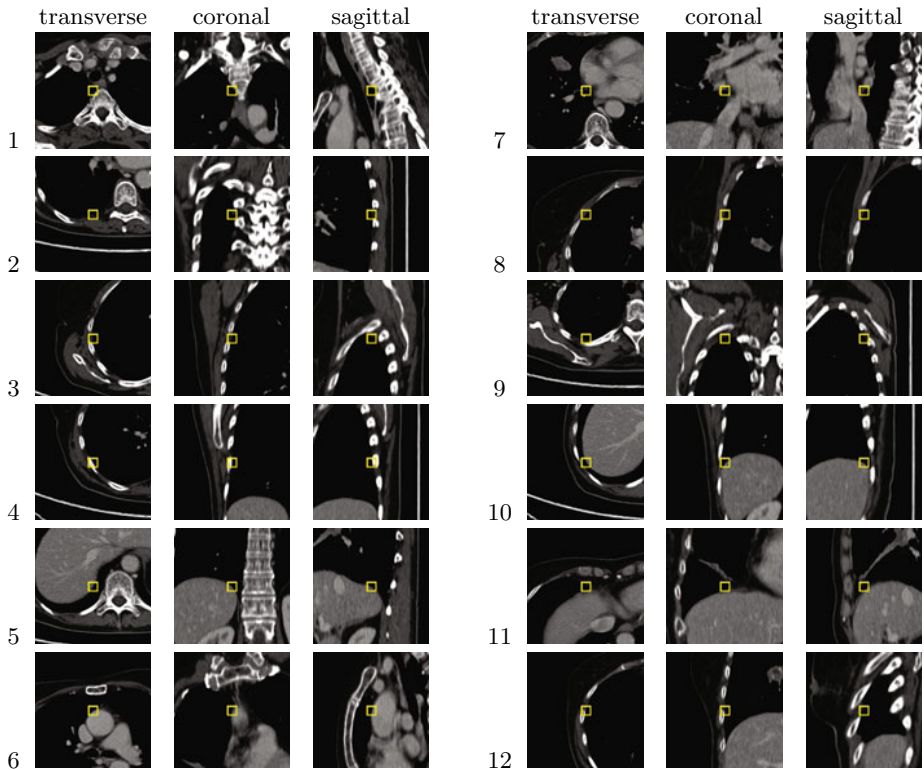


Fig. 4. Twelve strongest landmarks selected out of 614 with focus on spatial coverage. Notice that they are selected near distinctive anatomical structures such as ribs (3, 4, 5, 12), vertebrae (1, 2) and top (5) and bottom of the lung (9, 10, 11).

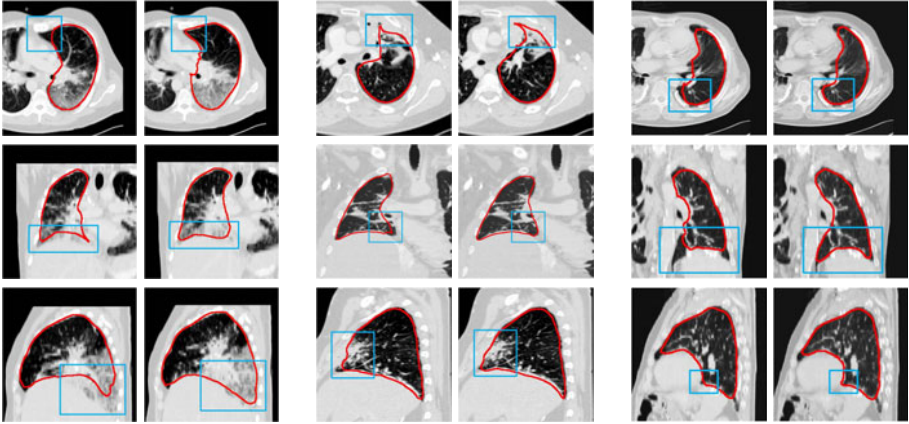


Fig. 5. Comparison of some results without stable landmarks (left column of every set) and with stable landmark detection (right column of every set)

trained using candidates formed from mesh vertices. The seemingly larger errors are caused by incorrect correspondences after mesh alignment. However, these landmarks are stable to provide accurate initialization for the mesh boundary refinement. We used 20 mm radius in landmark spatial filtering which resulted in 143 landmarks for the right lung and 133 landmarks for the left lung. To illustrate the effectiveness of the filtering, we set the radius to 70 mm to produce a set of 12 stable landmarks (Fig. 4). Notice that they are distributed across distinct locations inside the lung. Typically, landmarks near ribs and vertebrae are stable but not always. For example, landmark neighborhoods along some parts of the ribs or even across different ribs might not be distinctive enough.

Our final set of results analyzes performance of the lung segmentation. The algorithm was run as described in Section 2 and Figure 2. In the first experiment, the set of stable landmarks was used to initialize the boundary refinement. In the second experiment, the initialization was done by a mean mesh transformed according to the estimated pose. The errors summarized in Table 1 show that the initialization provided by stable landmarks helps to achieve significantly better accuracy ($p < 0.05$) of the final segmentations. The maximum error also decreased considerably and in the case of left lung one large failure was corrected. Several qualitative segmentation results involving pathologies are shown in Figure 5.

4 Conclusion

We proposed a robust learning-based technique for accurate lung segmentation in challenging CT volumes involving abnormalities. The technique first reliably detects the carina of the trachea as an anchor point for pose estimation of left and right lung. The poses are used to initialize a set of stable anatomical landmarks distributed on the lung surface. The stable landmarks are selected automatically

from a candidate set formed from vertices of mesh annotations by employing measures of uncertainty and spatial distribution. The initial landmark positions are refined and subsequently used to provide a rough estimate for the shape model and final lung boundary refinement.

We have shown the automatic landmark selection procedure determines a set of stable landmarks. These landmarks lead to improved initialization of the boundary refinement and ultimately higher accuracy of the final segmentations. Our future work focuses on further improvements especially near the lung sharp boundaries which are difficult to capture with a mesh representation.

References

1. De Nunzio, G., Tommasi, E., Agrusti, A., Cataldo, R., De Mitri, I., Favetta, M., Maglio, S., Massafra, A., Quarta, M., Torsello, M., et al.: Automatic lung segmentation in CT images with accurate handling of the hilar region. *Journal of Digital Imaging*, 1–17 (2009)
2. Hu, S., Hoffman, E., Reinhardt, J.: Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images. *IEEE Trans. Med. Imag.* 20(6), 490–498 (2002)
3. Kockelkorn, T., van Rikxoort, E., Grutters, J., van Ginneken, B.: Interactive lung segmentation in CT scans with severe abnormalities. In: *IEEE Int. Symp. on Biomed. Imag.*, pp. 564–567 (2010)
4. Korfiatis, P., Kalogeropoulou, C., Karahaliou, A., Kazantzi, A., Skiadopoulos, S., Costaridou, L.: Texture classification-based segmentation of lung affected by interstitial pneumonia in high-resolution CT. *Medical Physics* 35, 5290 (2008)
5. Ling, H., Zhou, S.K., Zheng, Y., Georgescu, B., Suehling, M., Comaniciu, D.: Hierarchical, learning-based automatic liver segmentation. In: *IEEE Conf. on Comp. Vis. and Patt. Recog.*, Los Alamitos, CA, USA (2008)
6. Prasad, M., Brown, M., Ahmad, S., Abtin, F., Allen, J., da Costa, I., Kim, H., McNitt-Gray, M., Goldin, J.: Automatic segmentation of lung parenchyma in the presence of diseases based on curvature of ribs. *Acad. Radiol.* 15(9), 1173–1180 (2008)
7. Sluimer, I., Prokop, M., van Ginneken, B.: Toward automated segmentation of the pathological lung in CT. *IEEE Trans. Med. Imag.* 24(8), 1025–1038 (2005)
8. Sofka, M., Zhang, J., Zhou, S., Comaniciu, D.: Multiple object detection by sequential Monte Carlo and hierarchical detection network. In: *IEEE Conf. on Comp. Vis. and Patt. Recog.* (June 13–18, 2010)
9. Sun, S., McLennan, G., Hoffman, E.A., Beichel, R.: Model-based segmentation of pathological lungs in volumetric CT data. In: *The Third International Workshop on Pulmonary Image Analysis* (2010)
10. Tu, Z.: Probabilistic boosting-tree: learning discriminative models for classification, recognition, and clustering. In: *IEEE Int. Conf. on Comp. Vis. and Patt. Recog.*, vol. 2, pp. 1589–1596 (2005)
11. Vidal, C., Hewitt, J., Davis, S., Younes, L., Jain, S., Jedynek, B.: Template registration with missing parts: Application to the segmentation of M. tuberculosis infected lungs. In: *IEEE Int. Symp. on Biomed. Imag.*, pp. 718–721 (2009)
12. Wang, J., Li, F., Li, Q.: Automated segmentation of lungs with severe interstitial lung disease in CT. *Medical Physics* 36, 4592 (2009)

Author Index

- Abolmaesumi, Purang I-65, I-379
Abugharbieh, Rafeef I-235, II-90,
II-285, III-603
Afsham, Narges I-65
Afshin, Mariam III-107
Ahmadi, Seyed-Ahmad I-73, III-362
Aichert, André I-73
Aja-Fernández, Santiago II-226
Akgul, Yusuf Sinan III-158
Alexander, Andrew L. II-217
Alexander, Daniel C. II-82
Ali, Sahirzeeshan I-661
Allain, Pascal I-500
Allard, Jérémie I-315
Andersson, Patrik II-549
André, Barbara III-297
Andrews, Shawn III-651
Angelini, Elsa D. I-387
Angelopoulou, Elli III-141
Armospach, Jean-Paul II-1
Arnold, Corey III-659
Arnold, Patrik II-623
Arridge, Simon I-581
Ashraf, Ahmed B. III-546
Asman, Andrew J. II-107
Assemlal, Haz-Edine II-157
Astner, Sabrina T. I-484
Atasoy, Selen III-83
Audière, Stéphane I-387
Aung, Tin III-1, III-91
Auzias, Guillaume II-310
Avni, Uri III-199
Awate, Suyash P. II-484
Axel, Leon I-468
Ayache, Nicholas I-500, II-663, III-297
Aylward, Stephen R. II-639
- Bagci, Ulas III-215
Baiker, Martin II-516
Baka, Nora II-434
Balter, James I-548
Bardana, Davide D. I-186
Bargiacchi, Anne II-9
Barkovich, A. James II-476
- Barratt, Dean I-605
Bartholmai, Brian J. III-223
Bartlett, Adam I-347
Bartsch, Ivonne III-454
Baskaran, Mani III-91
Basu, Anasuya I-1
Batmanghelich, Nematollah III-17
Bauer, Franz I-41
Bauer, Stefan III-354
Baust, Maximilian I-178, III-362
Beache, Garth M. III-587
Becker, Meike II-492, II-500
Beg, M. Faisal II-376
Bellani, Marcella II-426
Ben Ayed, Ismail III-107
Bergeest, Jan-Philip I-645
Bergeles, Christos I-33
Bériault, Silvain I-259
Bernasconi, Andrea II-352, III-445
Bernasconi, Neda II-352, III-445
Bernhardt, Dominik III-25
Bhushan, Manav I-476, II-541
Bi, Jinbo III-75
Bi, Xiaoming III-479
Biesdorf, Andreas I-589
Bigdelou, Ali I-129, I-178
Bilello, Michel II-532
Bilger, Alexandre I-339
Birkbeck, Neil III-338, III-667
Biros, George I-396, II-532
Bischof, Horst I-621
Blanc, Frédéric II-1
Blasco, Jordi I-355
Blauth, Michael II-393
Bloy, Luke II-234
Blumberg, Hilary P. II-33
Boctor, Emad M. I-371
Boese, Jan I-219, III-471, III-487
Boetzel, Kai III-362
Bogunović, Hrvoje III-330, III-395
Boone, Darren I-508
Booth, Brian G. II-90
Borgeat, Louis I-323
Bouget, David I-331

- Bouix, S. II-58
 Bou-Sleiman, Habib II-409
 Boussaid, Haithem III-346
 Bousse, Alexandre I-581
 Brady, J. Michael II-541
 Brambilla, Paolo II-426
 Brattain, Laura J. I-105
 Bricault, Ivan I-137
 Brieu, Nicolas III-579
 Brion, Véronique II-226
 Brost, Alexander I-540
 Brounstein, Anna I-235
 Buchner, Anna M. III-297
 Burdette, Everett Clif I-379, II-615
 Burgert, Oliver I-275
 Butakoff, Constantine II-50

 Caban, Jesus III-215
 Cachia, Arnaud II-9
 Caffo, Brian C. II-107
 Cai, Weidong III-191
 Campbell, Jennifer II-157
 Cao, Yihui III-272
 Capson, David W. I-460
 Cárdenes, Rubén III-330, III-395
 Cardoso, M. Jorge II-467, III-378
 Carranza-Herrezuelo, Noemí I-573
 Carrillo, Xavier III-411
 Caruyer, Emmanuel II-116
 Caserta, Enrico II-343
 Castellani, Umberto II-426
 Cathier, Pascal I-500
 Cattin, Philippe C. II-623
 Cavallaro, Alexander II-607
 Chaari, Lofti II-260
 Chai, Ping III-428
 Charbit, Maurice I-387
 Cheah, Andre I-637
 Chemouny, Stéphane II-508
 Chen, Hanbo II-318
 Chen, Mingqing III-471
 Chen, Terrence I-161, I-243, I-540
 Chen, Thomas Kuiran I-299
 Chen, Ting II-574, III-595
 Chen, Wufan III-570
 Chen, Xin II-680
 Chen, Xinjian III-387
 Cheng, Jack C.Y. II-384
 Cheng, Jian II-98
 Cheng, Jun III-1, III-91

 Cheng, Li I-637
 Cheshire, Nick I-49
 Cheung, Carol Y. III-1
 Chintalapani, Gouthami II-417
 Chng, Nick I-307
 Choti, Michael A. I-371
 Choyke, Peter L. III-272
 Christudass, Christhunesa I-661
 Chu, Winnie C.W. II-384
 Chui, C.K. I-428
 Chung, Albert C.S. III-436
 Chung, Moo K. II-217, II-302
 Cinquin, Philippe I-137, I-203
 Ciompi, Francesco III-411
 Ciuciu, Philippe II-9, II-260
 Clarkson, Matthew J. II-467
 Cobzas, Dana II-557
 Collins, D. Louis III-149
 Comaniciu, Dorin I-161, I-219, I-243,
 I-452, III-25, III-403, III-471, III-479,
 III-487, III-504
 Conant, Emily III-562
 Cotin, Stéphane I-315, I-339
 Coulon, Olivier II-310
 Coupé, Pierrick III-149
 Courtecuisse, Hadrien I-315
 Cox, B.T. I-363
 Crainiceanu, Ciprian II-107
 Criminisi, Antonio III-49, III-239,
 III-288
 Crowie, Lisa I-49
 Cui, Guangbin II-42
 Cvejic, Ana III-579

 Dalca, Adrian III-537
 Danagoulian, Giovanna III-537
 Daoud, Mohammad I-379
 Datar, Manasi II-368
 Datta, Saurabh III-512
 Davatzikos, Christos II-459, II-532,
 III-17
 Dawant, Benoit M. III-305
 Daye, Dania III-546
 de Bruijne, Marleen II-434
 De Camilli, Pietro I-629
 Declerck, Jérôme III-338, III-667
 De Craene, Mathieu III-256
 Deguet, Anton II-615
 Dehghan, Ehsan I-291, I-307, II-615
 de Jong, Pim A. III-207

- Delgado Leyton, Edgar J.F. III-9
 Delingette, Hervé I-500
 Delis, Foteini III-611
 del Nido, Pedro J. III-520
 Del Rio Barquero, Luís Miguel II-393
 De Luca, V. I-597
 Demirci, Stefanie I-178
 Deng, Zhigang I-25, I-251
 Denny, Kevin II-277
 Depeursinge, Adrien III-231
 Dequidt, Jérémie I-339
 de Raedt, Sepp II-360
 Deriche, Rachid II-98, II-116
 de Seze, Jérôme II-1
 Deshmukh, Sanchit I-291
 Dewan, Maneesh II-451, II-574
 Dhollander, Thijs II-166
 Díez, Jose Luis III-395
 Ding, Yu I-564
 Dohi, Takeyoshi I-81, I-169
 Dojat, Michel II-260
 Dong, Aoyan III-17
 Donner, Sabine III-454
 Dragon, Ralf III-454
 Duchateau, Josselin I-203
 Duchateau, Nicolas III-256
 Duchesnay, Edouard II-9
 Duffau, Hugues II-508
 Duncan, James S. I-629
 Duong, Christophe III-512
 Duriez, Christian I-315, I-339
 Durrleman, Stanley II-401
 Dyrby, Tim B. II-82
- Eavani, Harini II-234
 Eberl, Stefan III-191
 Ecabert, Olivier I-161
 Eckstein, Hans-Henning I-178
 Egan, Gary F. II-293
 Eisenschenk, Stephan J. III-595
 El-Baz, Ayman III-175, III-587
 El-Ghar, Mohamed Abou III-175, III-587
 Elhawary, Haytham I-211
 El-Kara, Leila III-659
 Ellisman, Mark I-670
 Elnakib, Ahmed III-175, III-587
 Epstein, Jonathan I. I-661
 Escalera, Sergio III-496
- Eskildsen, Simon F. III-149
 Etyngier, Patrick I-500
- Falk, Robert III-175
 Falk, Volkmar I-275
 Fallavollita, Pascal I-73, II-615
 Fan, Jianqing II-269
 Fan, Xiaoyao I-412
 Fang, Shiaofen II-376
 Fasel, Beat II-623
 Feldman, Michael III-546
 Feng, Dagan III-191
 Feng, Qianjin III-570
 Feuerstein, Marco II-582
 Fichtinger, Gabor I-291, I-299, I-307, II-615
 Fillard, Pierre II-9
 Fischer, Bernd I-436
 Fischer, Clara II-310
 Fishbaugh, James II-401
 Fitchinger, Gabor I-379
 Fleury, Gilles I-97
 Foncubierta-Rodríguez, Antonio III-231
 Fonov, Vladimir III-149
 Forbes, Florence II-260
 Fouard, Céline I-137
 Fouque, Anne-Laure II-9
 Framme, Carsten I-33
 Frangi, Alejandro F. I-355, II-50, II-393, III-256, III-330, III-395
 Freiman, M. II-74
 Friman, Ola I-436
 Frisoni, Giovanni B. II-663
 Fritsch, Virgile III-264
 Fritscher, Karl III-554
 Fritscher, Karl D. II-393
 Fu, Chi-Wing II-384
 Fu, Y.B. I-428
 Fuhrmann, Simon II-500
 Fulham, Michael J. III-191
 Funka-Lea, Gareth I-243, III-403, III-471
- Gangloff, Jacques I-57
 Gao, Fei I-492
 Gao, Mingchen I-468
 Gaonkar, Bilwaj II-459
 Gardiazabal, José II-582
 Gatta, Carlo III-411, III-496

- Gavenonis, Sara III-546
 Gavrilesco, Maria II-293
 Ge, Bao II-149
 Gee, James III-562
 Gehlbach, Peter I-1
 Georgescu, Bogdan I-219, I-452, III-479,
 III-504
 Gerber, Samuel II-484
 Gerig, Guido II-401
 Ghafaryasl, Babak II-50
 Gholipour, Ali II-124
 Ghosh, Aurobrata II-98
 Ghotbi, Reza I-178
 Gie, Robert III-133
 Gill, Raja II-327
 Gilmore, John H. II-66, II-133, III-313
 Gimel'farb, Georgy III-175, III-587
 Giovanello, Kelly II-269
 Giuly, Rick I-670
 Gleeson, Fergus V. II-541
 Glenn, Orit A. II-476
 Glocker, Ben III-239
 Godin, Guy I-323
 Goela, Aashish III-107
 Goldberger, Jacob III-199
 Golland, Polina III-537
 Gooya, Ali II-532
 Goussard, Pierre III-133
 Grabner, H. I-597
 Grady, Leo III-512
 Graf, Franz II-607
 Graham, Jim II-680
 Grasser, Andreas I-41
 Grbić, Saša I-219
 Greenspan, Hayit III-199
 Grigis, Antoine II-1
 Groher, Martin II-582, III-579
 Groves, Adrian R. II-647
 Gu, Xianfeng II-335, II-384
 Guehring, Jens I-564, III-479
 Guerrero, Ricardo II-566
 Guetter, Christoph I-564
 Guo, Lei II-42, II-149, II-251, II-318,
 III-635
 Gur, Yaniv II-368
 Guy, Pierre I-235

 Haar, Gail ter I-605
 Habas, Piotr A. II-476
 Hacihaliloglu, Ilker I-235
 Häfner, Michael III-280
 Hager, Gregory D. I-1, I-145, I-371
 Hagmann, Cornelia F. III-378
 Haidegger, Tamás III-619
 HajGhanbari, Bahareh III-651
 Halligan, Steve I-508
 Hamady, Mohamad I-49
 Hamarneh, Ghassan II-90, III-603,
 III-651
 Hamprecht, F.A. I-653
 Hampshire, Thomas I-508
 Hamrouni, Sameh I-524
 Han, Junwei II-149
 Hanaoka, Shouhei III-554
 Handa, Jim I-1
 Hansen, Torben B. II-360
 Harouni, Ahmed A. I-444
 Hart, Gabriel L. II-639
 Hashizume, Makoto I-169
 Hata, Nobuhiko I-41
 Hatt, Charles I-283
 Hautmann, Hubert I-17
 Hawkes, David I-508, I-605
 Hayashi, Naoto III-554
 Hayes, Carmel III-479
 He, Ying II-384
 Heffter, Tamas I-299
 Heibel, Hauke II-582
 Heinrich, Mattias P. I-476, II-541
 Heisterkamp, Alexander III-454
 Heitz, Fabrice II-1
 Hennemuth, Anja I-436
 Hernández Hoyos, Marcela III-9
 Hernández-Vela, Antonio III-496
 Heye, Tobias I-589
 Hirano, Yasushi III-183
 Hirsch, Sven I-404
 Ho, Harvey I-347
 Ho, Hon Pong II-33
 Hodgson, Antony I-235
 Hojjatoleslami, Ali II-25
 Holzhey, David I-275
 Honnorat, Nicolas I-9
 Hoogendoorn, Corné II-50
 Hornegger, Joachim I-540, III-471,
 III-504
 Hosch, Waldemar I-589
 Hosseinbor, Ameer Pasha II-217
 Houle, Helene III-504
 Howe, Robert D. I-105, III-520

- Hu, Jiayi II-335
Hu, Mingxing I-508
Hu, Xintao II-149, II-318
Hua, Jing II-335
Huang, Heng III-115
Huang, Junzhou II-451, III-611
Huang, Kun II-343
Huang, Xiaolei III-611
Huang, Yang II-655
Humbert, Ludovic II-393
Hungar, Nikolai I-137
Hunter, Peter I-347
Hurtig, Mark B. I-267
Hush, Don III-9
Hutchinson, Charles II-680
Hutton, Brian F. I-581
- Ieiri, Satoshi I-169
Iglesias, Juan Eugenio III-58, III-659
Igo, Stephen R. III-512
Igual, Laura III-496
Imani, Farhad I-379
Ingalhalikar, Madhura II-234
Inlow, Mark II-376
Inoue, Jiro I-267
Ionasec, Razvan Ioan I-219, I-452, III-504
Irfanoglu, M.O. II-174
Irimia, Andrei II-639
Irving, Benjamin III-133
Islam, Ali III-107
- Jacobs, Colin III-207
Jacobs, Michael A. I-444
Jagadeesh, Vignesh I-613
Jain, Ameet I-153, I-283
James Stewart, A. I-121, I-186, I-267
Jannin, Pierre I-331
Janoos, Firdaus II-343
Jedynak, Bruno I-1
Jenkinson, Mark I-476, II-541
Jeong, Won-Ki I-621
Ji, Songbai I-412
Jia, Hongjun III-643
Jiang, Jiayan III-58
Jiang, Tianzi II-98
Jiang, Yifeng III-528
John, Matthias I-219, I-275, III-487
Johnston, Leigh A. II-293
Jolly, Marie-Pierre I-564, III-479
- Juloski, Aleksandar Lj. III-141
Juluru, Krishna III-49
Jurrus, Elizabeth I-670
- Kaftan, Jens III-338, III-667
Kakadiaris, Ioannis A. I-396
Kammerlander, Christian II-393
Kang, Hyejin II-302
Karamalis, Athanasios III-362
Karar, Mohamed Esmail I-275
Karwoski, Ronald A. III-223
Kazhdan, Michael II-442
Keller, Brad III-562
Kellman, Peter III-479
Kelm, B. Michael III-25
Kendall, Giles S. III-378
Khalifa, Fahmi III-175, III-587
Kido, Shoji III-183
Kikinis, Ron III-537
Kim, Boong-Nyun II-302
Kim, Hosung II-352, III-445
Kim, Kio II-476
Kim, Minsuok I-355
Kim, Sungeun II-376
Kirschner, Matthias II-492, II-500
Kitasaka, Takayuki I-194, III-248
Klein, Stefan I-573, II-549
Klein, Tassilo III-362
Klinder, Tobias III-454
Kneser, Reinhard III-463
Knott, G. I-653
Kobayashi, Etsuko I-113, I-428
Koch, Martin I-540
Kohlberger, Timo III-338, III-667
Kontos, Despina III-546, III-562
Köthe, U. I-653
Kratochvil, Bradley E. I-33
Kriegel, Hans-Peter II-607
Krishna Nand, K. II-90
Kroeker, Randall III-479
Krüger, Alexander III-454
Krupa, Alexandre I-57
Kulp, Scott I-468
Kumar, Ritwik I-670
Kummer, Michael P. I-33
Kunz, Manuela I-267
Kuo, Nathanael II-615
Kutter, Oliver I-73, III-512
Kuwana, Kenta I-81, I-169
Kwitt, Roland III-280

- Ladikos, Alexander I-516
 Lahalle, Elisabeth I-97
 Lai, Rongjie II-327
 Lalys, Florent I-331
 Landman, Bennett A. II-107
 Langet, Hélène I-97
 Larrabide, Ignacio I-355, III-395
 Lasser, Tobias I-227
 Lasso, Andras I-299, I-379
 Lauritsch, Guenter III-471
 Lauzon, Carolyn B. II-107
 Law, Max W.K. III-107
 Lazenec, Jean-Yves III-346
 Le Bihan, Denis II-226
 Lee, Beng Hai III-91
 Lee, Dong Soo II-302
 Lee, Hyekyoung II-302
 Lee, Junghoon II-615
 Lee, Su-Lin I-49
 Lefèvre, Julien II-310
 Le Floch, Edith II-9
 Lehotsky, Ákos III-619
 Lekadir, Karim II-50
 Lelieveldt, Boudewijn P.F. II-434,
 II-516
 Leone, Gustavo II-343
 Le Troter, Arnaud II-310
 Li, Gang II-671
 Li, John M. I-186
 Li, Kaiming II-42, II-251, II-318
 Li, Shuo III-107
 Li, Wei III-570
 Li, Xiang II-251
 Li, Xiuli III-387
 Li, Xuelong III-272
 Lian, Jun I-532
 Liang, Liang I-629
 Liao, Hongen I-81, I-113
 Liao, Rui I-540
 Liao, Shu III-570
 Licegevic, Oleg III-141
 Lichtman, Jeff I-621
 Lim, Chulwoo II-251
 Lin, Shi II-384
 Lin, Stephen III-1
 Lin, Weili II-66, II-133, II-269, III-313
 Lin, Xiang I-113
 Little, Stephen H. III-512
 Liu, Cheng-Yi III-58
 Liu, David III-166
 Liu, Huafeng I-420, I-492
 Liu, Jiang III-1, III-91
 Liu, Meizhu III-41, III-75
 Liu, Shizhen III-512
 Liu, Tianming II-42, II-149, II-251,
 II-318
 Liu, Zhao III-124
 Lobo, Julio I-291, I-307
 Lorenzi, Marco II-663
 Low, Adrian III-428
 Lówik, Clemens W.G.M. II-516
 Lu, Le III-41, III-75
 Lu, Xiaoguang III-479
 Lucas, Blake C. II-442
 Lueth, Tim C. I-41
 Lui, Lok Ming II-384
 Luo, Xiongbiao I-17, I-194, III-248
 Lv, Jinglei II-149

 Maas, Mario II-360
 Machiraju, Raghu II-174, II-343
 Macho, Juan M. I-355
 Macq, Benoît II-590
 Madabhushi, Anant I-661, III-33
 Maddah, Mahnaz II-191
 Maes, Frederik II-166
 Mahapatra, Dwarikanath III-420
 Maier, Thomas I-41
 Mangin, Jean-François II-226
 Manjón, José V. III-149
 Manjunath, B.S. I-613
 Manning, Samantha III-587
 Mansi, Tommaso I-219, I-452, II-352,
 III-445, III-504
 Marami, Bahram I-460
 Marlow, Neil III-378
 Martin-Yuste, Victoria III-496
 Marx, Gerald R. III-520
 Masamune, Ken I-81, I-169
 Massicotte, Philippe I-323
 Masutani, Yoshitaka III-554
 Mateus, Diana III-83, III-239
 Matin, Tahreema II-541
 Mauri, Josepa III-411
 McClure, Patrick III-587
 Meer, Peter III-25
 Meijering, Erik I-573
 Meining, Alexander III-83
 Melbourne, Andrew III-378
 Mendizabal-Ruiz, E. Gerardo I-396

- Mengue, Etienne Assoumou I-452, III-504
- Menzel, Manuela I-17
- Merlet, Sylvain II-116
- Merrifield, Robert III-627
- Metaxas, Dimitris N. I-468, II-451, III-611
- Metz, Coert II-434
- Mewes, Philip W. III-141
- Michael Brady, J. I-476
- Michailovich, O. II-58
- Mies, Carolyn III-546
- Miette, Véronique I-387
- Miller, James V. II-191
- Mirtuono, Pasquale II-426
- Mittal, Sushil III-25
- Modat, Marc II-467, III-378
- Moffitt, Richard A. III-66
- Mohr, David II-327
- Mohr, Friedrich-Wilhelm I-275
- Mok, Kelvin I-259
- Möller, Axel Martinez III-239
- Mollura, Daniel J. III-215
- Mollus, Sabine III-463
- Moradi, Mehdi I-291, I-307
- Morales, Hernán G. I-355
- Mori, Kensaku I-17, I-194, III-248
- Morris, William J. I-291, I-307
- Mountney, Peter I-89
- Mousavi, Parvin I-379
- Mueller, Edgar III-479
- Mueller, Kerstin III-471
- Muir, Lindsay II-680
- Mulkern, R.V. II-74
- Müller, Henning III-231
- Müller, Oliver III-454
- Müller, Tobias I-589
- Mung, Jay I-153
- Muñoz-Moreno, Emma II-50
- Murino, Vittorio II-426
- Nadeau, Caroline I-57
- Nagy, Melinda III-619
- Najafi, Mohammad I-65
- Nakajima, Susumu I-81
- Nakazawa, Toji I-113
- Nathan, Diane III-562
- Navab, Nassir I-17, I-73, I-129, I-178, I-219, I-227, I-484, II-582, III-83, III-239, III-362, III-579
- Navkar, Nikhil V. I-25, I-251
- Nekolla, Stephan III-239
- Nelson, Bradley J. I-33
- Nemoto, Mitsutaka III-554
- Neumann, Dominik III-141
- Ng, Bernard II-285
- Nho, Kwangsik II-376
- Nie, Feiping III-115
- Nie, Jingxin II-671, III-635
- Niessen, Wiro I-573, II-434
- Niethammer, Marc II-639, II-655
- Nitzken, Matthew III-175
- Noack, T. I-452
- Noble, Jack H. III-305
- Noblet, Vincent II-1
- Nolte, Lutz-Peter II-409, III-354
- Nüsslin, Fridtjof I-484
- Ohdaira, Takeshi I-169
- Ohtomo, Kuni III-554
- Oktay, Ayse Betul III-158
- Onceanu, Dumitru I-121
- Ong, Sim Heng III-428
- Orkisz, Maciej III-9
- Osman, Nael F. I-444
- Ostermeier, Martin I-161
- Ourselin, Sebastien I-581, II-467, III-378
- Ouseph, Rosemary III-587
- Ouwehand, Willem III-579
- Pace, Danielle F. II-639
- Padoy, Nicolas I-145
- Palmore, Tara N. III-215
- Paniagua, Beatriz II-368
- Papademetris, Xenophon II-33, III-528
- Paragios, Nikos I-9, I-97, II-508, III-346
- Parisot, Sarah II-508
- Parry, R. Mitchell III-66
- Parthasarathy, Vijay I-283
- Pathak, Sayan III-49
- Paulsen, Keith D. I-412
- Pauly, Olivier III-239
- Pécot, Thierry II-343
- Pedemonte, Stefano I-581
- Peikari, Mohammad I-299
- Pelletier, Daniel II-327
- Pengcheng, Shi II-242
- Pennec, Xavier II-631, II-663
- Perez-Rossello, J.M. II-74

- Perrin, Douglas P. III-520
 Perrot, Matthieu II-310
 Peters, Jochen III-463
 Peters, Terry III-107
 Petrović, Aleksandar II-524
 Petrusca, L. I-597
 Pfefferbaum, Adolf II-191
 Pfister, Hanspeter I-621, I-670
 Piella, Gemma III-256
 Pierpaoli, C. II-174
 Pike, G. Bruce I-259, II-157
 Plate, Annika III-362
 Pluim, Josien P.W. II-549
 Pohl, Kilian M. II-459, II-532
 Poirier, Guillaume I-323
 Poline, Jean-Baptiste II-285, III-264
 Pölsterl, Sebastian II-607
 Pontabry, Julien II-209
 Popovic, Aleksandra I-211
 Poupon, Cyril II-226
 Poupon, Fabrice II-226
 Pozo, José María III-330
 Prakosa, Adityo I-500
 Preiswerk, Frank II-623
 Prêteux, François I-524
 Prince, Jerry L. I-556, II-615
 Prummer, Simone I-161
 Pujol, Oriol III-411
 Punithakumar, Kumaradevan III-107
 Punwani, Shonit I-508

 Qian, Zhen I-468

 Radeva, Petia III-411, III-496
 Raghunath, Sushravya III-223
 Rajagopalan, Srinivasan III-223
 Rajagopalan, Vidya II-476
 Ramachandran, Bharat I-211
 Rambaldelli, Gianluca II-426
 Rangarajan, Anand III-595
 Rasiwasia, Nikhil III-280
 Rathi, Yogesh II-58
 Rathke, Fabian III-370
 Ratnarajah, Nagulan II-25
 Raval, Amish I-283
 Raykar, Vikas III-75
 Raza, S. Hussain III-66
 Régis, Jean II-310
 Reiber, Johan H.C. II-516
 Reichl, Tobias I-17

 Reid, W. Darlene III-651
 Reyes, Mauricio II-409, II-631, III-354
 Richa, Rogerio I-1
 Riddell, Cyril I-97
 Riff, Olivier II-226
 Riffaud, Laurent I-331
 Riga, Celia I-49
 Rijkhorst, Erik-Jan I-605
 Risacher, Shannon L. II-376, III-115
 Risholm, Petter I-548
 Risser, Laurent I-476
 Ritacco, Lucas E. II-409
 Rittscher, Jens II-343
 Rivaz, Hassan I-371
 Rivens, Ian I-605
 Robb, Richard A. III-223
 Robert, Adeline I-137
 Roberts, David W. I-412
 Roberts, Mike I-621
 Roberts, Timothy P.L. II-234
 Robertson, Nicola J. III-378
 Rohkohl, Christopher III-471
 Rohlfing, Torsten II-191
 Rohling, Robert I-65
 Rohr, Karl I-589, I-645
 Rosen, Mark III-546
 Rosenhahn, Bodo III-454
 Ross, Ian III-107
 Roth, Holger I-508
 Roth, Tobias II-393
 Rougon, Nicolas I-524
 Rousseau, François II-209, II-476
 Rueckert, Daniel II-566

 Saalbach, Axel III-463
 Sabuncu, Mert R. III-99
 Sadikot, Abbas F. I-259
 Sakuma, Ichiro I-113
 Salcudean, Septimiu E. I-291, I-307
 Salganicoff, Marcos III-41, III-75
 Salomir, R. I-597
 Salomir, Rares II-623
 Saloux, Eric I-500
 Sammet, S. II-174
 San Roman, Luis I-355
 Sánchez, Clara I. III-207
 Sara Mahdavi, S. I-291
 Saur, Stefan C. III-207
 Saykin, Andrew J. II-376, III-115
 Schaap, Michiel II-434

- Scheffler, Klaus II-623
 Scherrer, Benoit II-124
 Schmidt, Ehud III-537
 Schmidt, Michaela III-479
 Schmidt, Stefan III-370
 Schnabel, Julia A. I-476, II-541, II-647
 Schneider, Matthias I-404
 Schneider, Robert J. III-520
 Schnörr, Christoph III-370
 Schubert, Matthias II-607
 Schubert, Rainer II-393, III-554
 Schultz, Robert T. II-234
 Schultz, Thomas II-141
 Schwarz, Loren Arthur I-129
 Schwenke, Michael I-436
 Scott, Julia II-476
 Seeburger, J. I-452
 Seiler, Christof II-631
 Sen, Abhishek II-557
 Serbanovic-Canic, Jovana III-579
 Sermesant, Maxime I-500
 Seshamani, Sharmishta II-417
 Setsompop, K. II-58
 Seydhosseini, Mojtaba I-670
 Shah, Dipan J. I-25, I-251
 Shen, Dinggang I-532, II-17, II-66,
 II-133, II-200, II-277, II-671, III-313,
 III-570, III-635, III-643
 Shen, Hongying I-629
 Shen, Li II-376, III-115
 Shen, Tian III-611
 Shenton, M.E. II-58
 Shi, Feng III-313, III-635
 Shi, Kuangyu I-484
 Shi, Pengcheng I-420, I-492
 Shi, Yonggang II-327
 Sicotte, Nancy II-327
 Siddiqi, Kaleem II-157
 Sijbers, Jan II-166
 Simmons, Andy II-25
 Simon, Johannes III-141
 Simon, Loic III-346
 Simonetti, Orlando P. I-564
 Simonyan, Karen III-288
 Simpson, Ivor J.A. II-647
 Singh, Shantanu II-343
 Sinusas, Albert J. III-528
 Sirouspour, Shahin I-460
 Slabaugh, Greg I-508
 Smal, Ihor I-573
 Smith, Lyndon III-124
 Smith, Melvyn III-124
 Sofka, Michal III-166, III-338, III-667
 Sokhadze, Guela III-587
 Song, Danny II-615
 Song, Yang III-191
 Sonmez, Ahmet E. I-25
 Soza, Grzegorz III-166
 Speier, Peter III-479
 Speier, William III-659
 Staib, Lawrence H. II-33, III-528
 Stankovic, Zoran I-283
 Stanton, Douglas I-211
 Staring, Marius II-516, II-549
 Stilling, Maiken II-360
 Stone, Maureen I-556
 Stoyanov, Danaïl I-89
 Straehle, C.N. I-653
 Strauss, Gero I-41
 Streekstra, Geert J. II-360
 Strobel, Norbert I-540
 Studholme, Colin II-476
 Styner, Martin II-368
 Subaie, Fahd Al I-259
 Suenaga, Hideyuki I-81
 Suetens, Paul II-166
 Suffredini, Anthony F. III-215
 Sühling, Michael III-166
 Sullivan, Edith V. II-191
 Sun, Jiuai III-124
 Sun, Liang I-484
 Sun, Ying III-420, III-428
 Sunaert, Stefan II-166
 Swaminathan, Shanker II-376
 Székely, Gábor I-404, I-597
 Szilágyi, László III-619
 Sznitman, Raphael I-1
 Tachibana, Rie III-183
 Tanner, C. I-597
 Tansella, Michele II-426
 Tao, Dacheng III-91
 Taquet, Maxime II-590
 Tasdizen, Tolga I-670
 Taskar, Ben III-17
 Taylor, Paul III-133
 Taylor, Russell H. I-1, II-417, II-442
 Tek, Huseyin III-403
 Tenenhaus, Arthur I-97
 Tenenholtz, Neil A. III-520

- Teo, C.L. I-428
 Thanos, Panayotis III-611
 Thienphrapa, Paul I-211
 Thirion, Bertrand II-285, III-264
 Thyreau, Benjamin II-9, III-264
 Tian, Jie III-387
 Tibrewal, Radhika I-178
 Tietjen, Christian III-166
 Todd-Pokropek, Andrew III-133
 Toga, Arthur W. II-327
 Tomikawa, Morimasa I-169
 Toomre, Derek K. I-629
 Top, Andrew III-603
 Totz, Johannes I-89
 Tran, Huy Hoang I-81
 Treeby, Bradley E. I-363
 Tristán-Vega, Antonio II-182, II-226
 Troussel, Yves I-97
 Tsekos, Nikolaos V. I-25, I-251
 Tsymbal, Alexey III-25
 Tu, Zhuowen III-58, III-659
 Tumen, Mustafa I-363
 Turkbey, Baris III-272
 Twellmann, Thorsten III-207

 Uhl, Andreas III-280
 Unger, Markus I-621
 Uzunbas, Mustafa III-611

 Vaillant, Régis I-9
 van de Giessen, Martijn II-360
 Van de Ville, Dimitri III-231
 van Ginneken, Bram III-207
 Van Hecke, Wim II-166
 Van Horn, John D. II-639
 Van Leemput, Koen III-99
 Vannan, Mani III-512
 van Vliet, Lucas J. II-360
 Varoquaux, Gael II-285, III-264
 Vasconcelos, Nuno III-280
 Vaupel, Peter I-484
 Vázquez-Reina, Amelio I-621
 Vega-Higuera, Fernando III-25, III-403
 Veltri, Robert I-661
 Vemuri, Baba C. III-595
 Veraart, Jelle II-166
 Vercauteren, Tom III-297
 Verma, Ragini II-234
 Vespa, Paul M. II-639
 Vetter, Christoph I-227

 Vialard, François-Xavier II-655
 Vignon, Francois I-153
 Villain, Nicolas I-500
 Villa-Uriol, Maria-Cruz I-355
 Vincent, Thomas II-260
 Voigt, Ingmar I-452, III-504
 Volkow, Nora III-611
 von Tengg-Kobligk, Hendrik I-589
 Voros, Sandrine I-203
 Voros, Szilard I-468
 Vos, Frans M. II-360
 Voss, S.D. II-74
 Vu, Nhat I-613

 Wachinger, Christian I-178
 Wächter-Stehle, Irina III-463
 Waldman, Stephen D. I-267
 Walker, L. II-174
 Wallace, Michael B. III-297
 Wan, Jing II-376
 Wang, Chaohui III-346
 Wang, Defeng II-384
 Wang, Fei II-33
 Wang, Hua III-115
 Wang, Jiaping II-269
 Wang, Lei II-376
 Wang, Lejing I-73, I-178
 Wang, Li III-313
 Wang, Lichao III-627
 Wang, Lihong II-277
 Wang, Linwei I-420
 Wang, May D. III-66
 Wang, Peng I-161
 Wang, Qian I-532, II-17, II-200
 Wang, Tianzhou III-487
 Wang, Yan III-562
 Wang, Yang I-219
 Wang, Yaping III-635
 Warfield, Simon K. II-74, II-124, II-590,
 III-322
 Warr, Robert III-124
 Weber, Bruno I-404
 Weber, Tim Frederik I-589
 Wee, Chong-Yaw II-277
 Weese, Jürgen III-463
 Wei, Dong III-428
 Wein, Wolfgang I-516, III-512
 Weiner, Michael W. II-376
 Weisenfeld, Neil I. III-322
 Welk, Martin III-554

- Wells, William M. I-548
Wendler, Thomas I-227
Wesarg, Stefan II-492, II-500
Westin, Carl-Fredrik II-58, II-182
Wetzl, Jens III-338, III-667
Whitaker, Ross II-368, II-484
Whitmarsh, Tristan II-393
Wieczorek, Matthias I-73
Wilkens, Jan J. I-484
Wimmer, Andreas I-540
Witte, Frank III-454
Wolf, Matthias III-75
Wolf, Rémi I-203
Wolz, Robin II-566
Wong, Damon Wing Kee III-91
Wong, Ken C.L. I-420
Wong, Tien Yin III-1, III-91
Woo, Jonghye I-556
Woolrich, Mark W. II-647
Wörz, Stefan I-589
Wrba, Friedrich III-280
Wu, Dijia III-166
Wu, Guorong I-532, II-17, II-200,
III-643
Wu, Mark Z. I-379
Wu, Wen I-540
Wu, Yu-Chien II-217
- Xiao, Gaoyu III-33
Xin, Shi-Qing II-384
Xu, Dong III-1
Xu, Rui III-183
Xu, Yanwu III-1
Xue, Hui I-564
- Yan, Pingkun III-272
Yang, Guang-Zhong I-49, I-89, III-83,
III-627
Yao, Jianhua III-215, III-387
- Yap, Pew-Thian II-17, II-66, II-133,
II-200, II-277, III-635
Yasuda, Ryuya I-113
Yazdanpanah, Azadeh III-651
Ye, Ning I-637
Ye, Xiaojing III-41
Yeniaras, Erol I-25, I-251
Young, Andrew N. III-66
Yu, Shipeng III-41
Yu, Weimiao I-637
Yuan, Yuan III-272
- Zhan, Yiqiang II-451, II-574
Zhang, Daoqiang II-277, III-643
Zhang, Hui II-82
Zhang, Jingdan III-338, III-667
Zhang, Shaoting I-468, II-451, II-574,
III-611
Zhang, Tuo II-42, II-149, II-318
Zhang, Xing III-387
Zheng, Guoyan II-598
Zheng, Yefeng I-219, III-25, III-403,
III-471, III-487
Zheng, Yuanjie III-562
Zhenghui, Hu II-242
Zhou, S. Kevin III-25, III-166, III-338,
III-403, III-487, III-667
Zhou, Xiang Sean II-451
Zhu, Dajing II-42
Zhu, Hongtu II-269
Zhu, Ning III-436
Zhu, Peihong II-484
Zhuang, Zhen W. III-528
Zilbovicius, Monica II-9
Zisserman, Andrew III-288
Zöllei, Lilla II-524
Zou, Guangyu II-335
Zuehlsdorff, Sven I-564
Zuluaga, Maria A. III-9
Zuo, Siyang I-169