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

Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012

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

3 Part III

MICCAI

Springer

Lecture Notes in Computer Science 7512

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

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

Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012

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

Springer

Volume Editors

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

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

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

Preface

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

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

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

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

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

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

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

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

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

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

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

October 2012

Nicholas Ayache
Hervé Delingette
Polina Golland
Kensaku Mori

Accepted MICCAI 2012 Papers

by Clinical Theme

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

by Technical Theme

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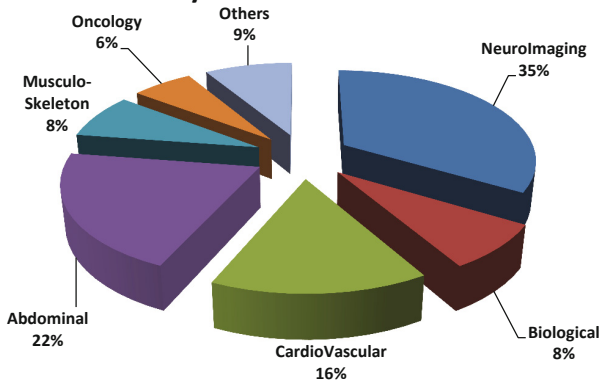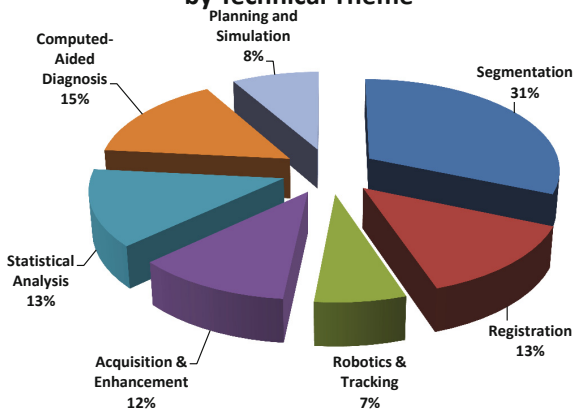
by Country of First Author

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

Organization

General Chair

Nicholas Ayache   Inria, Sophia Antipolis, France

Program Chair and Co-chairs

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

Workshops, Tutorials and Challenges Chair and Co-chairs

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

MICCAI Society, Board of Directors

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

Consultants to Board

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

Program Committee

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

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

Organizing Institution

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

Local Organizing Committee

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

Liaison with the MICCAI Society

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

Official Partners

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

Sponsors

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

Exhibitors

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

Reviewers

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

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

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

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

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

Essert, Caroline
Fahrig, Rebecca
Fallavollita, Pascal
Fan, Yong
Farag, Aly
Fedorov, Andriy
Fei, Baowei
Felblinger, Jacques
Fenster, Aaron
Fetita, Catalin
Fiebich, Martin
Figl, Michael
Fischer, Gregory
Fishbaugh, James
Fitzpatrick, J. Michael
Fleig, Oliver
Florack, Luc
Fonov, Vladimir
Foroughi, Pezhman
Fouard, Céline
Fradkin, Maxim
Freiman, Moti
Friboulet, Denis
Fripp, Jurgen
Fritzsche, Klaus H.
Frouin, Frédérique
Frouin, Vincent
Funka-Lea, Gareth
Fuster, Andrea
Gagnon, Langis
Gangloff, Jacques
Ganz, Melanie
Gao, Mingchen
Gao, Wei
Gao, Yi
Garcia-Lorenzo, Daniel
Garvin, Mona
Gassert, Roger
Gatenby, Chris
Gee, Andrew
Georgescu, Bogdan
Georgii, Joachim
Geremia, Ezequiel
Ghanbari, Yasser
Gholipour, Ali

Ghosh, Aurobrata
Giannarou, Stamatia
Gibaud, Bernard
Gibson, Eli
Gilles, Benjamin
Gilson, Wesley
Giusti, Alessandro
Glaunès, Joan Alexis
Glocker, Ben
Gobbi, David
Goh, Alvina
Goksel, Orcun
Gonzalez Ballester, Miguel Angel
González Osorio, Fabio Augusto
Gooding, Mark
Goodlett, Casey
Gorges, Sebastien
Graham, Jim
Gramfort, Alexandre
Grass, Michael
Grau, Vicente
Grenier, Thomas
Griswold, Mark
Guerrero, Julian
Guetter, Christoph
Guevara, Pamela
Gulsun, Mehmet Akif
Gur, Yaniv
Gutman, Boris
Hacihaliloglu, Ilker
Hahn, Horst
Hajnal, Joseph
Hall, Timothy
Hamarneh, Ghassan
Hanahusa, Akihiko
Hanaoka, Shouhei
Hans, Arne
Hansen, Michael Sass
Hanson, Dennis
Hao, Xiang
Hartov, Alexander
Hastreiter, Peter
Hatt, Chuck
Haynor, David
He, Huiguang

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Awards Presented at MICCAI 2011, Toronto

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

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

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

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

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

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

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

Table of Contents – Part III

Diffusion Imaging: From Acquisition to Tractography

Image Acquisition, Segmentation and Recognition

Image Registration II

NeuroImage Analysis II

Analysis of Microscopic and Optical Images II

Image Segmentation III

Diffusion Weighted Imaging II

Computer-Aided Diagnosis and Planning II

Microscopic Image Analysis

Accelerated Diffusion Spectrum Imaging
with Compressed Sensing Using Adaptive Dictionaries

Berkin Bilgic[1], Kawin Setsompop[2,3], Julien Cohen-Adad[2,3], Van Wedeen[2,3], Lawrence L. Wald[2,3,4], and Elfar Adalsteinsson[1,4]

[1] Massachusetts Institute of Technology, MA, USA
[2] A.A. Martinos Center for Biomedical Imaging, MA, USA
[3] Harvard Medical School, MA, USA
[4] Harvard-MIT Division of Health Sciences and Technology, MA, USA

Abstract. Diffusion Spectrum Imaging (DSI) offers detailed information on complex distributions of intravoxel fiber orientations at the expense of extremely long imaging times (~1 hour). It is possible to accelerate DSI by sub-Nyquist sampling of the q-space followed by nonlinear reconstruction to estimate the diffusion probability density functions (pdfs). Recent work by Menzel et al. imposed sparsity constraints on the pdfs under wavelet and Total Variation (TV) transforms. As the performance of Compressed Sensing (CS) reconstruction depends strongly on the level of sparsity in the selected transform space, a dictionary specifically tailored for sparse representation of diffusion pdfs can yield higher fidelity results. To our knowledge, this work is the first application of adaptive dictionaries in DSI, whereby we reduce the scan time of whole brain DSI acquisition from 50 to 17 min while retaining high image quality. *In vivo* experiments were conducted with the novel 3T *Connectome* MRI, whose strong gradients are particularly suited for DSI. The RMSE from the proposed reconstruction is up to 2 times lower than that of Menzel et al.'s method, and is actually comparable to that of the fully-sampled 50 minute scan. Further, we demonstrate that a dictionary trained using pdfs from a single slice of a particular subject generalizes well to other slices from the same subject, as well as to slices from another subject.

1 Introduction

Diffusion weighted MR imaging is a widely used method to study the interconnectivity and structure of the brain. Diffusion Tensor Imaging (DTI) is an established diffusion weighted imaging method, which models the diffusion as a univariate Gaussian distribution (1). One limitation of this model arises in the presence of fiber crossings, and this can be addressed by using a more involved imaging method. Diffusion Spectrum Imaging (DSI) samples the full q-space and yields a complete description of the diffusion probability density function (pdf) (2). While DSI is capable of resolving complex distributions of intravoxel fiber orientations, full q-space coverage comes at the expense of substantially long scan times (~1 hour).

Compressed Sensing (CS) comprises algorithms that recover data from undersampled acquisitions by imposing sparsity or compressibility assumptions on the reconstructed

images (3). In the domain of DSI, acceleration with CS was successfully demonstrated by Menzel et al. (4) by imposing wavelet and Total Variation (TV) penalties in the pdf space. Up to an undersampling factor of 4 in *q*-space, it was reported that essential diffusion properties such as orientation distribution function (odf), diffusion coefficient, and kurtosis were preserved (4).

The performance of CS recovery depends strongly on the level of sparsity of the signal in the selected transform domain (3). While prespecified transformations such as wavelets and spatial gradients yield sparse signal representation, tailoring the sparsifying transform based on the characteristics of the particular signal type may offer even sparser results. K-SVD is an algorithm that designs a dictionary that achieves maximally sparse representation of the input training data (5). The benefit of using data-driven, adaptive dictionaries trained with K-SVD was also demonstrated in CS reconstruction of structural MR imaging (6).

In this work, we employ the K-SVD algorithm to design a sparsifying transform that captures the structure in diffusion pdfs and yields a signal representation with increased level of sparsity. Coupling this adaptive dictionary with the FOcal Underdetermined System Solver (FOCUSS) algorithm (7), we obtain a parameter-free CS algorithm. With 3-fold undersampling of *q*-space, we demonstrate *in vivo* up to 2-fold reduced pdf reconstruction errors relative to our implementation of the CS algorithm that uses wavelets and variational penalties by Menzel et al. (4). At higher acceleration factors of 5 and 9, we still demonstrate substantial improvement. For additional validation, the RMSE of the reconstructed 'missing' diffusion images were calculated by comparing them to a gold standard dataset obtained with 10 averages. In this case, dictionary-based reconstructions were seen to be comparable to the fully-sampled 1 average data. Further, we show that a dictionary trained on data from a particular subject generalizes well to reconstruction of another subject's data, still yielding significantly reduced reconstruction errors. Hence, application of the proposed method might reduce a typical 50-minute DSI scan to 17 minutes (upon 3× acceleration) while retaining high image quality. Additionally, we also investigate using a simple ℓ_1-norm penalty in the pdf space with the FOCUSS algorithm, and show that this approach gives comparable results to the more involved wavelet- and TV-based reconstruction by Menzel et al. (4), while being computationally more efficient.

2   Theory

CS Recovery with Prespecified Transforms

Letting $\boldsymbol{p} \in \mathbb{C}^N$ represent the 3-dimensional diffusion pdf at a particular voxel as a column vector, and $\boldsymbol{q} \in \mathbb{C}^M$ denote the corresponding undersampled *q*-space information, CS recovery with wavelet and TV penalties aim to solve the convex optimization problem at a single voxel,

$$min_{\boldsymbol{p}} \ \|\mathbf{F}_\Omega \boldsymbol{p} - \boldsymbol{q}\|_2^2 + \alpha \cdot \|\mathbf{\Psi}\boldsymbol{p}\|_1 + \beta \cdot \text{TV}(\boldsymbol{p}) \qquad \text{(Eq. 1)}$$

where $\mathbf{F}_\Omega$ is the undersampled Fourier transform operator, $\mathbf{\Psi}$ is a wavelet transform operator, $\text{TV}(.)$ is the Total Variation penalty, and α and β are regularization parameters that need to be determined.

Training an Adaptive Transform with K-SVD

Given an ensemble $\mathbf{P} \in \mathbb{C}^{N \times L}$ formed by concatenating L example pdfs $\{\mathbf{p}_i\}_{i=1}^{L}$ collected from a training dataset as column vectors, the K-SVD algorithm (5) aims to find the best possible dictionary for the sparse representation of this dataset by solving,

$$min_{\mathbf{P,D}} \sum_{i=1}^{L} \|\mathbf{x}_i\|_0 \text{ subject to } \|\mathbf{P} - \mathbf{DX}\|_F^2 \leq \varepsilon \qquad \text{(Eq. 2)}$$

where $\mathbf{X}$ is the matrix that contains the transform coefficient vectors $\{\mathbf{x}_i\}_{i=1}^{L}$ as its columns, $\mathbf{D}$ is the adaptive dictionary formed by example pdfs, ε is a fixed constant that adjusts the data fidelity, and $\|.\|_F$ is the Frobenius norm. The K-SVD works iteratively, first by fixing $\mathbf{D}$ and finding an optimally sparse $\mathbf{X}$ using orthogonal matching pursuit, then by updating each column of $\mathbf{D}$ and the transform coefficients corresponding to this column to increase data consistency.

CS Recovery with an Adaptive Transform Using FOCUSS

The FOCUSS algorithm aims to find a sparse solution to the underdetermined linear system $\mathbf{F}_\Omega \mathbf{D} \mathbf{x} = \mathbf{q}$, where $\mathbf{x}$ is the vector of transform coefficients in the transform space defined by the dictionary $\mathbf{D}$ using the following iterations,

For iteration number $t = 1, ... T$,

$$W_{j,j}^t = diag\left(\left|x_j^t\right|^{1/2}\right) \qquad \text{(Eq. 3)}$$

$$\mathbf{s}^t = \text{argmin}_s \|\mathbf{s}\|_2^2 \text{ such that } \mathbf{F}_\Omega \mathbf{D} \mathbf{W}^t \mathbf{s} = \mathbf{q} \qquad \text{(Eq. 4)}$$

$$\mathbf{x}^{t+1} = \mathbf{W}^t \mathbf{s}^t \qquad \text{(Eq. 5)}$$

Here, $\mathbf{W}^t$ is a diagonal weighting matrix whose jth diagonal entry is denoted as $W_{j,j}^t$, $\mathbf{x}^t$ is the estimate of transform coefficients at iteration t whose jth entry is x_j^t. The final reconstruction in diffusion pdf space is obtained via the mapping $\mathbf{p} = \mathbf{D} \mathbf{x}^{T+1}$.

We note that it is possible to impose sparsity-inducing ℓ_1 penalty directly on the pdf coefficients by taking $\mathbf{D}$ to be the identity matrix $\mathbf{I}$. A detailed description of application of FOCUSS algorithm to MRI can be found in (8), where it is shown that reweighted ℓ_2 norm solutions for the auxiliary variable $\mathbf{s}^t$ induce ℓ_1 penalty on $\mathbf{x}^t$.

3   Methods

Diffusion EPI acquisitions were obtained from three healthy volunteers (subjects A, B and C) using a novel 3T system (Magnetom Skyra Connectom, Siemens Healthcare, Erlangen, Germany) equipped with the AS302 "Connectom" gradient with G_{max} = 300 mT/m (here we used G_{max} = 200 mT/m) and Slew = 200 T/m/s. A custom-built 64-channel RF head array (9) was used for reception with imaging parameters of 2.3 mm isotropic voxel size, FOV = 220×220×130, matrix size = 96×96×57, b_{max} = 8000 s/mm^2, 514 directions full sphere q-space sampling organized in Cartesian grid with interspersed b=0 image every 20 TRs (for motion correction), in-plane acceleration = 2× (using GRAPPA algorithm), TR/TE = 5.4 s / 60 ms, total imaging time ~50 min. In addition, at

5 q-space points ($[1,1,0]$, $[0,2,-1]$, $[0,0,3]$, $[0,4,0]$, and $[5,0,0]$) residing on 5 different shells, 10 averages were collected for noise quantification. Eddy current related distortions were corrected using the reversed polarity method (10). Motion correction (using interspersed b=0) was performed using FLIRT (11) with sinc interpolation.

Variable-density undersampling (using a power-law density function (3)) with $R = 3$ acceleration was applied in q-space on a 12×12×12 grid. Three different adaptive dictionaries were trained with data from slice 30 of subjects A, B and C. Reconstruction experiments were applied on test slices that are different than the training slices. In particular, two reconstruction experiments were performed. First, voxels in slice 40 of subject A were retrospectively undersampled in q-space, and reconstructed using 5 different methods: wavelet+TV method of Menzel et al. (4), ℓ_1-regularized FOCUSS, and Dictionary-FOCUSS with the three dictionaries trained on three different subjects. Second, voxels in slice 25 of subject B were undersampled with the same $R = 3$ sampling pattern, and again reconstructed with wavelet+TV, ℓ_1-FOCUSS, and the three dictionaries trained on three different subjects. For Menzel et al.'s method, Haar wavelets in MATLAB's wavelet toolbox were used. The regularization parameters α and β in Eq.1 were chosen by parameter sweeping with values $\{10^{-4}, 3 \cdot 10^{-4}, 10^{-3}, 3 \cdot 10^{-3}\}$ to minimize the reconstruction error of 100 randomly selected voxels in slice 40 of subject A. The optimal regularization parameters were found to be $\alpha = 3 \cdot 10^{-4}$ for wavelet and $\beta = 10^{-4}$ for the TV term. By taking the fully-sampled data as ground-truth, the fidelity of the five methods were compared using root-mean-square error (RMSE) normalized by the ℓ_2-norm of ground-truth as the error metric both in pdf domain and q-space.

Since the fully-sampled data are corrupted by noise, computing RMSEs relative to them will include contributions from both reconstruction errors and additive noise. To address this, the additional 10 average data acquired at the selected 5 q-space points were used. As a single average full-brain DSI scan takes ~50 min, it was not practical to collect 10 averages for all of the undersampled q-space points. As such, we rely on both error metrics, namely: the RMSE relative to one average fully-sampled dataset and the RMSE relative to gold standard data for 5 q-space points.

4   Results

Fig.1 depicts the error of the five reconstruction methods in the pdf domain for each voxel in slice 40 of subject A. At $R = 3$ acceleration, reconstruction error of Menzel et al.'s method averaged over brain voxels in the slice was 15.8%, while the error was 15.0% for ℓ_1-regularized FOCUSS. Adaptive dictionary trained on subject A yielded 7.8% error. Similarly, reconstruction with dictionaries trained on pdfs of the other subjects B and C returned 7.8% and 8.2% RMSE, respectively. At $R = 5$ and 9, Dictionary-FOCUSS with training on subjects C and B returned 9.3% and 10.0% RMSE, respectively.

In Fig.2, reconstruction errors at $R = 3$ on slice 25 of subject B are presented. In this case, Menzel et al.'s method yielded 17.5% average RMSE, and ℓ_1-FOCUSS had 17.3% error. Dictionary trained on slice 40 of subject B returned 11.4% RMSE, while adaptive transforms trained on subjects A and C had 11.4% and 11.8% error, respectively. At higher acceleration factors of $R = 5$ and 9, Dictionary-FOCUSS with training on subjects C and A returned 13.5% and 14.2% RMSE, respectively.
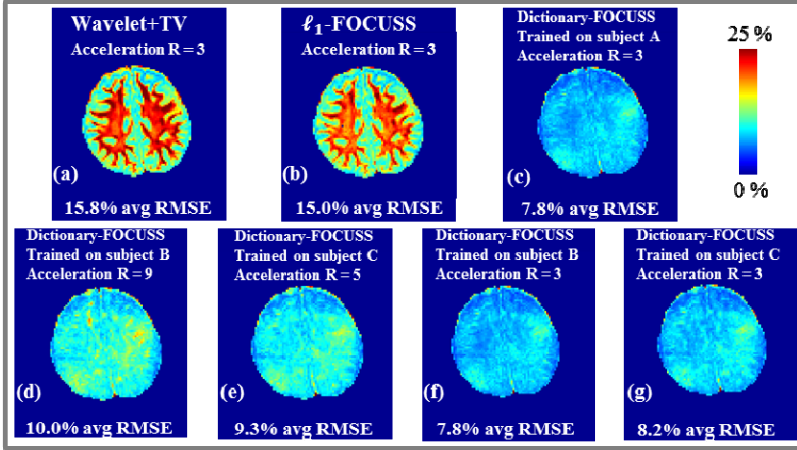
Fig. 1. RMSE at each voxel in slice 40 of subject A upon $R=3$ acceleration and reconstruction with Menzel et al.'s method (a), ℓ_1-FOCUSS (b), Dictionary-FOCUSS trained on subjects A (c), B (f), and C (g). Dictionary-FOCUSS errors in (d) and (e) are obtained at higher acceleration factors of $R=9$ and 5.
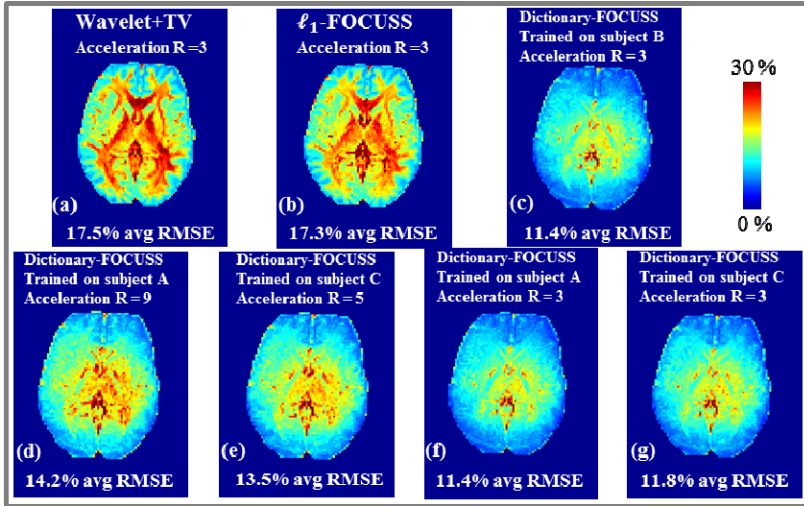
Fig. 2. RMSE at each voxel in slice 25 of subject B upon $R=3$ acceleration and reconstruction with Menzel et al.'s method (a), ℓ_1-FOCUSS (b), Dictionary-FOCUSS trained on subjects B (c), A (f), and C (g). Dictionary-FOCUSS errors in (d) and (e) are obtained at higher acceleration factors of $R=9$ and 5.

Reconstruction errors in q-space images of subject A obtained with Wavelet+TV, ℓ_1-FOCUSS and Dictionary-FOCUSS for the undersampled q-space directions are plotted in Fig.3. For two particular diffusion directions, q-space reconstructions obtained with the three methods are also presented.

To allow inter-subject comparison, slices that correspond to approximately the same anatomical region in subjects A, B and C were also reconstructed. As shown in Fig.1f, slice 40 of subject A when reconstructed using the dictionary trained on subject B gave 7.8% average error. Slices 38 of subjects B and C yielded 7.3% and 9.3% RMSE when reconstructed with the dictionary trained on subject A upon 3-fold undersampling (not shown).

In an attempt to quantify the noise in q-space and separate it from CS reconstruction error, we take the 10 average data acquired at 5 q-space directions as ground truth and compute RMSEs relative to them. Fig.4 shows the error plots for the 1 average fully sampled data, Wavelet+TV, ℓ_1-FOCUSS, and Dictionary-FOCUSS reconstructions relative to the 10 average data for slices from subjects A and B.
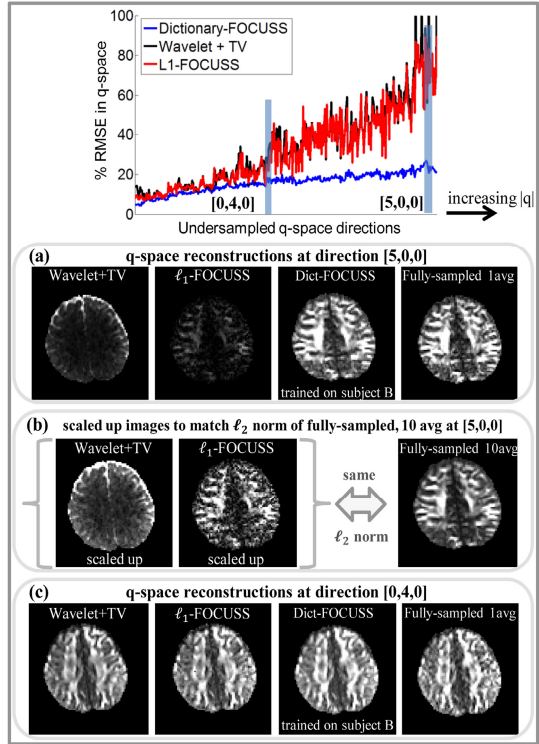
Fig. 3. Top: RMSEs in 'missing' q-space directions estimated with Wavelet+TV, ℓ_1 - and Dictionary-FOCUSS at R=3. q-space images at directions [5,0,0] (a) and [0,4,0] (c) are also depicted. In (b), Wavelet+TV and ℓ_1 recons of dir. [5,0,0] are scaled up to have same ℓ_2 norm as fully-sampled 10 avg .
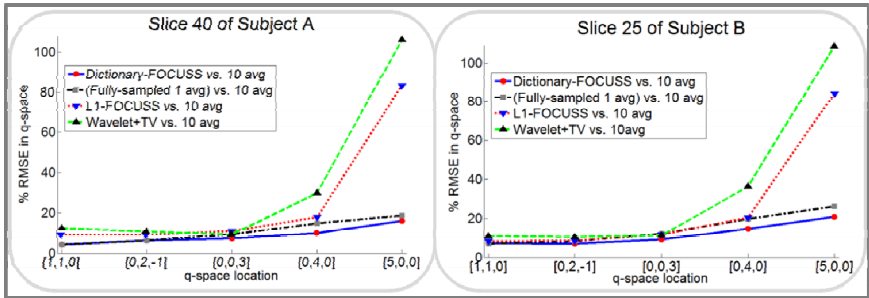
Fig. 4. Panel on the left depicts RMSEs of Wavelet+TV, ℓ_1-FOCUSS and Dictionary-FOCUSS at R=3 and fully-sampled 1 average data computed in 5 q-space locations relative to the 10 average data for subject A. Panel on the right shows the same comparison for the slice belonging to subject B.

Fig.5 presents Orientation Distribution Function (ODF) visualization of reconstructions obtained at 3-fold acceleration using Wavelet+TV and Dictionary-FOCUSS, and compares the tractography solutions obtained with adaptive reconstruction and fully-sampled data.
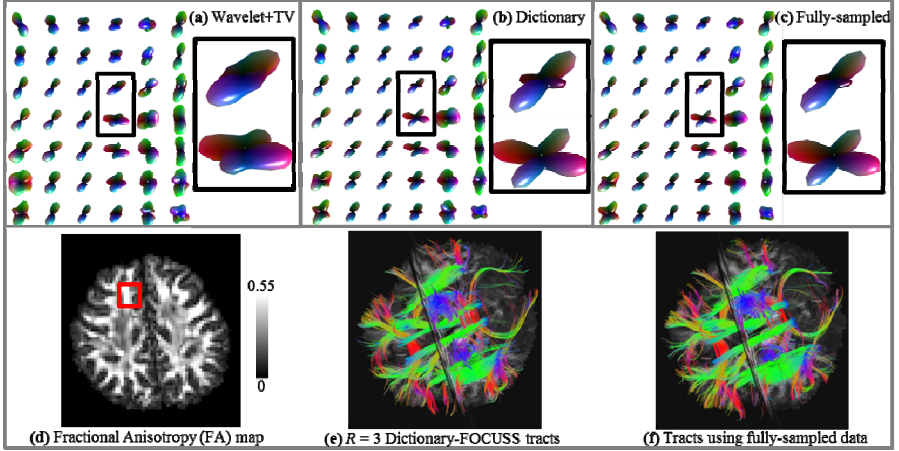
Fig. 5. ODFs for subject A using Wavelet+TV (a), dictionary (b), and fully-sampled data (c) within the ROI in the FA map in (d). Tracts with $R=3$ dictionary recon (e) and fully-sampled data (f) are also presented.

5   Discussion

This work presented the first application of adaptive transforms to voxel-by-voxel CS reconstruction of undersampled q-space data. Relative to reconstruction with prespecified transforms, the proposed algorithm has up to 2 times reduced error in the pdf domain at the same acceleration factor ($R = 3$), while requiring no regularization parameter tuning. When the undersampling ratio was increased to $R = 5$ and even up to $R = 9$, the proposed method still demonstrated substantial improvement relative to using prespecified transforms at $R = 3$ (Figs.1 and 2). This will render DSI clinically feasible, by cutting a 50 min scan to 5.5 min upon 9-fold acceleration. As demonstrated, a dictionary trained with pdfs from a single slice of a particular subject generalizes to other slices of the same subject, as well as to different subjects. However, further tests are needed to see if dictionaries can generalize across healthy and patient populations, across age groups, or fundamentally different anatomical locations.

Since the acquired 1 average DSI data is corrupted by noise (especially in the outer shells), it is desired to obtain noise-free data for more reliable computation of CS reconstruction errors. Because even the 1 average full-shell acquisition takes ~50 min, it is practically not possible to collect multiple-average data at all q-space points. To address this, one representative q-space sample at each shell was collected with 10 averages to serve as "(approximately) noise-free" data. When the noise-free data were

taken to be ground-truth, the dictionary reconstruction with 3-fold undersampling was comparable to the fully-sampled 1 average data for both subjects (Fig.4).

RMSE in Fig.2 was overall higher than in Fig.1. A possible explanation is the inherently lower signal-noise-ratio (SNR) in the lower axial slice, particularly in the center area of the brain which is further away from the receive coils. In particular, the error is higher in the central region of the image where the SNR is expected to be lowest. Future work includes a detailed analysis of how SNR level affects the reconstruction performance of the proposed CS algorithm and the dictionary learning step.

As seen in Fig.3, wavelet and TV penalized reconstruction and ℓ_1-FOCUSS yield especially poor quality results in estimating the high q-space samples. In particular, as depicted in Fig.3a, these CS methods tend to underestimate the high q-space content. However, this is not a simple scaling problem, even when q-space images are scaled to have the same ℓ_2-norm as the fully-sampled 10 average data, they yield either flat (Wavelet+TV) or grainy (ℓ_1-FOCUSS) results (Fig.3b). ODF visualization and tractography solutions in Fig.5 show good agreement between the adaptive reconstruction and the fully-sampled dataset. Average FA and tract volume metrics obtained from 18 major white-matter pathways were seen to support this good agreement (data not shown due to space limitation). As the tract results show that main fiber bundles are not corrupted, adaptive reconstruction causes no regional bias in the reconstruction.

In our implementation, *per voxel* processing time of ℓ_1-FOCUSS was 0.6 seconds, while this was 12 seconds for Dictionary-FOCUSS and 27 seconds for Wavelet+TV method on a workstation with 12GB memory and 6 processors. Hence, full-brain reconstruction using the Dictionary-FOCUSS algorithm would still take *several days*. Because each voxel can be processed independently, parallel implementation is likely to be a significant source of performance gain.

The proposed CS acquisition/reconstruction can be combined with other techniques to further reduce the acquisition time and/or improve reconstruction quality. In particular, combining the proposed method with the Blipped-CAIPI Simultaneous MultiSlice (SMS) acquisition (12) could reduce a 50 minute DSI scan to mere 5.5 minutes upon 9-fold acceleration (3×3 CS-SMS) while retaining high image quality.

References

1. Basser, P.J., Mattiello, J., LeBihan, D.: MR diffusion tensor spectroscopy and imaging. Biophys. J. 66(1), 259–267 (1994)
2. Wedeen, V.J., et al.: Mapping complex tissue architecture with diffusion spectrum magnetic resonance imaging. Magn. Reson. Med. 54(6), 1377–1386 (2005)
3. Lustig, M., Donoho, D., Pauly, J.M.: Sparse MRI: The application of compressed sensing for rapid MR imaging. Magn. Reson. Med. 58(6), 1182–1195 (2007)
4. Menzel, M.I., et al.: Accelerated diffusion spectrum imaging in the human brain using compressed sensing. Magn. Reson. Med. 66(5), 1226–1233 (2011)
5. Aharon, M., Elad, M., Bruckstein, A.: K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. IEEE T. Signal Proces. 54(11), 4311–4322 (2006)

6. Ravishankar, S., Bresler, Y.: MR image reconstruction from highly undersampled k-space data by dictionary learning. IEEE Trans. Med. Imaging 30(5), 1028–1041 (2011)
7. Gorodnitsky, I.F., Rao, B.D.: Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm. IEEE T. Signal Proces. 45(3), 600–616 (1997)
8. Ye, J.C., Tak, S., Han, Y., Park, H.W.: Projection reconstruction MR imaging using FOCUSS. Magn. Reson. Med. 57(4), 764–775 (2007)
9. Keil, B., et al.: A 64-channel brain array coil for 3T imaging. In: 20th Annual ISMRM Scientific Meeting and Exhibition (2012)
10. Bodammer, N., et al.: Eddy current correction in diffusion-weighted imaging using pairs of images acquired with opposite diffusion gradient polarity. Magn. Reson. Med. 51(1), 188–193 (2004)
11. Jenkinson, M., et al.: Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage 17(2), 825–841 (2002)
12. Setsompop, K., et al.: Blipped-controlled aliasing in parallel imaging for simultaneous multislice Echo Planar Imaging with reduced g-factor penalty. Magn. Reson. Med. (2011)

Parametric Dictionary Learning for Modeling EAP and ODF in Diffusion MRI

Sylvain Merlet, Emmanuel Caruyer, and Rachid Deriche

Athena Project-Team, INRIA Sophia Antipolis - Méditerranée, France

Abstract. In this work, we propose an original and efficient approach to exploit the ability of Compressed Sensing (CS) to recover Diffusion MRI (dMRI) signals from a limited number of samples while efficiently recovering important diffusion features such as the Ensemble Average Propagator (EAP) and the Orientation Distribution Function (ODF). Some attempts to sparsely represent the diffusion signal have already been performed. However and contrarily to what has been presented in CS dMRI, in this work we propose and advocate the use of a well adapted learned dictionary and show that it leads to a sparser signal estimation as well as to an efficient reconstruction of very important diffusion features. We first propose to learn and design a sparse and parametric dictionary from a set of training diffusion data. Then, we propose a framework to analytically estimate in closed form two important diffusion features : the EAP and the ODF. Various experiments on synthetic, phantom and human brain data have been carried out and promising results with reduced number of atoms have been obtained on diffusion signal reconstruction, thus illustrating the added value of our method over state-of-the-art SHORE and SPF based approaches.

1 Introduction

Diffusion MRI (dMRI) modality is known to assess the integrity of brain anatomical connectivity and to be very useful for examining and quantifying white matter microstructure and organization not available with other imaging modalities. However, dMRI data acquisition is also well known to be significantly time-consuming, in particular when Diffusion Spectrum Imaging (DSI) or High Angular Resolution Diffusion Imaging (HARDI) is to be used in a clinical setting. Accelerated acquisitions, relying on a smaller number of sampling points, are thus more than welcome. Compressed Sensing (CS) [3] is a recent technique to accurately reconstruct sparse signals from few measurements. In this work, we present a CS based method for accelerating the reconstruction of the EAP and the ODF, by significantly reducing the number of measurements. Some approaches have been recently proposed in order to build dictionaries that enable sparse representations (For a summary see [1]). However, these works lead to non-parametric dictionaries, which do not enable to obtain continuous representations of the diffusion signal neither allow to get closed form for diffusion features such as EAP and ODF. For instance, in [9] and in [10], the authors

nicely proposed a dictionary for a sparse modeling in dMRI. However, their dictionary is just postulated to be the Spherical Ridgelets in [9] and the Spherical Wavelets in [10] i.e the dictionary is not learnt from a training phase as ours. In addition, their dictionary is used only for modelling diffusion signal in [9] and only the ODF in [10] i.e not the EAP and the ODF.

In this work, we propose to learn a parametric dictionary based on a framework especially designed for dMRI. This framework enables a continuous modeling of the diffusion signal and leads to analytical and closed form formulas to estimate two important diffusion features : the EAP, which represents the full 3D displacement probability function of water molecules at every voxel and the ODF, which characterizes the relative likelihood of water diffusion along any given angular direction. The article is structured as follows : we start by introducing the dMRI framework together with the proposed basis, then we focus on the parametric dictionary learning algorithm and finally we conclude with an experimental part illustrating the added-value of our approach with promising results showing how our approach allows to accurately estimate the diffusion signal with much less atoms (almost the half) than using state-of-the-art bases such as SHORE [8,4] and SPF [2].

2 dMRI Framework for Recovery of EAP and ODF

In this section, we introduce the dMRI framework to model the diffusion signal and its important features: the EAP and the ODF. Due to a lack of space, we omit all the details and refer the interested reader to our research report, to be included in the final version of this article.

2.1 Basis for Diffusion Signal Estimation

Inspired by the basis proposed in the state-of-the-art diffusion signal estimation i.e the SHORE [8,4], the SPF [2], we propose to express the diffusion signal as a truncated linear combination of 3D functions Ψ_{nlm} where each basis function can be decomposed in a radial part and an angular part represented by a the symmetric and real spherical harmonics (SH) $Y(\mathbf{u})$ [5].

$$E(q\mathbf{u}) = \sum_{n=0}^{N} \sum_{l=0}^{L} \sum_{m=-l}^{l} c_{nlm} \Psi_{nlm}(\mathbf{a_{nlm}}, \zeta_{nlm}, q\mathbf{u}), \qquad (1)$$

with $c_{nlm} = \langle E, \Psi_{nlm} \rangle$ are the transform coefficients, N the radial order, L the angular order , q the norm of the effective gradient, $\mathbf{u}$ a unitary vector. We let $j = j(n, \ell, m) = (n+1) \times ((\ell^2 + \ell + 2)/2 + m)$ and rewrite formula 1 as $E(q\mathbf{u}) = \sum_{j=0}^{J} c_j \Psi_j(\mathbf{a_j}, \zeta_j, q\mathbf{u})$. Then, Ψ_j is given by

$$\Psi_j(\mathbf{a_j}, \zeta_j, q\mathbf{u}) = \frac{1}{\sqrt{\chi_{\mathbf{a_j}}}} \exp\left(-\frac{q^2}{2\zeta_j}\right) \left(\frac{q^2}{\zeta_j}\right)^{l/2} \sum_{i=0}^{n} \mathbf{a_j}(\mathbf{i}) \left(\frac{q^2}{\zeta_j}\right)^{i} Y_j(\mathbf{u}), \qquad (2)$$

where $\chi_{\mathbf{a_j}}$ is the normalization factor in the l_2-norm sense, $\mathbf{a_j} \in \mathbb{R}^{n+1}$ a vector of polynomial parameters associated with the scaled parameter ζ_j, and $Y_j = Y_\ell^m$ the spherical harmonic of order ℓ and degree m with ℓ, m the index related to j. $\mathbf{a_j}$ and ζ_j are the parameters we want to learn for each 3D function Ψ_j.

It's worthwhile to note that our basis Ψ_j in Eq. 2 simplifies to the SHORE basis [8,4] by setting $\mathbf{a_j}$ as the generalized Laguerre polynomial coefficients of degree n and l.

2.2 On EAP and ODF Recovery: Closed Formulas

Using the basis function we proposed in the previous section, it is possible to derive important and analytical closed formulas for the two diffusion features : the EAP and the ODF.

EAP Feature: The EAP, denoted $P(R\mathbf{r})$, is the inverse Fourier transform of the normalized diffusion signal, denoted $E(q\mathbf{u})$. From eq. 1, we can derive the following expression for the EAP :

$$P(\mathbf{a_j}, \zeta_j, R\mathbf{r}) = \sum_{j=0}^{J} c_j \frac{2\pi(-1)^{\frac{l}{2}}}{\sqrt{\chi_{\mathbf{a_j}} R}} Y_j(\mathbf{r}) \sum_{i=0}^{n} \frac{\mathbf{a_j(i)}}{\zeta_j^{\frac{l}{2}+i}} \frac{(2\pi R)^{l+\frac{1}{2}} \Gamma(\frac{3}{2}+l+i)}{2^{l+\frac{3}{2}}(\frac{1}{2\zeta_j})^{\frac{3}{2}+l+i}\Gamma(l+\frac{3}{2})} {}_1F_1(\frac{3}{2}+l+i; l+\frac{3}{2}; -2\zeta_j\pi^2 R^2),$$

(3)

where $_1F_1$ is the confluent hypergeometric function and Γ the Gamma function.

ODF Feature: The ODF is given by $\Upsilon(\mathbf{r}) = \int_0^\infty P(R\cdot\mathbf{r})R^2 dR$. We can show that the following closed form can be derived for the ODF :

$$\Upsilon((\mathbf{a_j}, \zeta_j, \mathbf{r}) = \frac{1}{\sqrt{4\pi}} + \sum_{j=0}^{J} Y_j(\mathbf{r}) \left(\sum_{n=0}^{N} c_j \frac{(-1)^{\frac{l}{2}}\pi^{l+1}}{2^{\frac{3}{2}}\sqrt{\chi_{\mathbf{a_j}}}} \left(\frac{1}{2\zeta_j\pi^2}\right)^{\frac{3}{2}+\frac{l}{2}} \sum_{i=0}^{n} \frac{\mathbf{a_j(i)}}{\zeta_j^{\frac{l}{2}+i}} \frac{\Gamma(\frac{l}{2}+\frac{3}{2})\Gamma(\frac{l}{2}+i)}{\left(\frac{1}{2\zeta_j}\right)^{\frac{3}{2}+l+i}\Gamma(\frac{l}{2})} \right).$$

(4)

3 Dictionnary Learning

Here, we introduce a parametric dictionary learning method that enables a sparse representation of any diffusion signal from continuous functions. We started by considering the K-SVD [1] algorithm as a model for our own method. However, the K-SVD technique designs non-parametric dictionaries, which presents some shortcomings among which : 1) a non-parametric method does not enable to compute a continuous version of our signal (not suitable for interpolation, neither data extrapolation) , 2) we could not get closed form for diffusion features, which would be very appreciated for EAP and ODF estimations, and 3) the K-SVD is acquisition-dependant. Although the K-SVD method appears powerful in designing sparse dictionary, these drawbacks push towards a better design via parametric dictionary learning. This algorithm consists in a sparse coding step and a dictionary update step, where the polynomial and scale parameters $\mathbf{a_j}$, ζ_j

are estimated using a non linear approach, the Levenberg-Marquardt algorithm (LMA). The section 3.1 presents our dictionary learning algorithm and the section 3.2 describes the method we use to reconstruct any diffusion signal using the dictionary previously learned.

3.1   Dictionary Learning Algorithm

Notation: Suppose the training set consists in M observations $\{\mathbf{s}_i\}_{i=1}^{M}$ (i.e. M voxels). For each observation s_i we have m_s samples in the q-space, i.e. $\mathbf{s}_{1..M} \in \mathcal{R}^{m_s}$. We represent $\{\mathbf{s}_i\}_{i=1}^{M}$ in matrix form $\mathbf{S} \in \mathcal{R}^{m_s \times M}$ where $\mathbf{s}_i$ is the ith column. The algorithm searches for the dictionary $\mathbf{D} \in \mathcal{R}^{m_s \times J}$, that enables the sparsest representation for every column of $\mathbf{S}$. The dictionary consists in J atoms $\{\mathbf{d}_j\}_{j=1}^{J}$ with $\mathbf{d}_j \in \mathcal{R}^{m_s}$ a column of $\mathbf{D}$. $\mathbf{d}_j$ corresponds to the 3D function Ψ_j in eq. 2. Here, we do not try to directly estimate $\mathbf{d}_j$ but the polynomial and scale parameters $\mathbf{a_j}$ and ζ_j, that characterize the atom $\mathbf{d}_j$. For each observation $\mathbf{s}_i$, we define a coefficient vector $\mathbf{c}_i \in \mathcal{R}^{n_c}$, which forms the ith column of the coefficient matrix $\mathbf{C} \in \mathcal{R}^{n_c \times M}$.

Given a training data set $\mathbf{S}$, we search for the dictionary D that gives the sparsest representation of this set. The overall problem is to find the dictionary $\mathbf{D}$ and the vectors $\mathbf{c}_i$ in $\mathbf{C}$ by solving :

$$\arg\min_{\mathbf{c}_i, \mathbf{D}} \{\|\mathbf{S} - \mathbf{DC}\|_2^2\} \;\; subject \;\; to \;\; \forall_i \|\mathbf{c}_i\|_1 \leq \epsilon \tag{5}$$

with $\epsilon \in \mathbb{R}$. The method to solve Eq. 5 is described in the following and a summary of the algorithm is given in alg. 1. This algorithm iteratively alternates between sparse signal estimations (i.e. $\mathbf{C}$) and updates of the dictionary (i.e. $\mathbf{D}$) so to better fit the training data set (i.e. $\mathbf{S}$).

In the first step, the estimation of the column vector $\mathbf{c}_i$ is performed separately for each signal $\mathbf{s}_i$, i.e for each column of $\mathbf{S}$. Sparse estimation is achieved through a fast iterative thresholding shrinkage algorithm (FISTA) [6].

In the second step, we update the dictionary $\mathbf{D}$. For this purpose, we compute an absolute averaged coefficient vector, $\hat{\mathbf{c}} = 1/M \sum_i |\mathbf{c}_i|$, and find the atoms associated with the non zeros values of $\hat{\mathbf{c}}$. It gives a rough idea of which atoms are used for modeling the signal and enables to discard some unnecessary atoms, which enforces sparsity. Then, in this set of atoms, we update one atom at a time, while fixing all the others. This process is repeated for all the non-zero coefficients in $\hat{\mathbf{c}}$. The in-update atom is denoted $\mathbf{d}_k$. To update this atom, we begin by decomposing the error term in eq. 5 as in [1], i.e.

$$\|\mathbf{S} - \mathbf{DC}\|_{\ell 2}^2 = \left\| \mathbf{S} - \sum_{j=1}^{M} \mathbf{d}_j \mathbf{c}_j^r \right\|_{\ell 2}^2 = \left\| \left(S - \sum_{j \neq k} \mathbf{d}_j \mathbf{c}_j^r \right) - \mathbf{d}_k \mathbf{c}_k^r \right\|_{\ell 2}^2 = \|\mathbf{E}_k - \mathbf{d}_k \mathbf{c}_k^r\|_{\ell 2}^2, \quad (6)$$

where $\mathbf{c}_j^r$ is the jth row of C. We could use the LMA in order to fit the atom $\mathbf{d}_k$ to the error matrix $\mathbf{E}_k$. However, because we take into account all the coefficients and atoms, this dictionary update doesn't enforce sparsity. Hence, we need to

---

Algorithm 1. Semi-parametric dictionary learning

1. Initialize the dictionary by fixing the polynomial and scale parameters $\mathbf{a_j}$, ζ_j.
2. Sparse estimation of the observations $\{\mathbf{s}_i\}_{i=1}^M$. We use the FISTA algorithm to solve for $\mathbf{c_i}$ associated to each observation : $\arg\min_{\mathbf{c_i}}\{\|\mathbf{s}_i - \mathbf{Dc}_i\|_2^2\}$ *subject to* $\forall_i\|\mathbf{c}_i\|_1 \leq \epsilon$ (Eq. 5bis).
3. Updating the dictionary. Compute the absolute averaged coefficients vector $\hat{\mathbf{c}} = 1/M \sum_i |\mathbf{c}_i|$. Repeat until all the atoms of the dictionary , with non zeros value in $\hat{\mathbf{c}}$, have been scanned :
   - Let note the current atom, the kth.
   - Define the group of observation that use this atom : $w_k = \{i, 1 \leq i \leq M, c_k(i) \neq 0\}$.
   - Compute the error matrix $\mathbf{E}_{w_k} \in \mathcal{R}^{m \times card(w_k)}$.
   - Apply Levenberg - Marquardt algorithm to estimate the polynomials and scale parameters $a_{k,1..n}$, ζ_k, which enable $\mathbf{d}_k$ to best fit $\mathbf{E}_{w_k}$
   - Update the atoms according to $a_{k,1..n}$, ζ_k.

4. Go back to the step 2 unless the overall error does not vary anymore

---

reduce the number of atoms used for modeling the signal. For this purpose, we define the group of observations that use the atom $\mathbf{d}_k$, i.e $w_k = \{i, 1 \leq i \leq M, c_i(k) \neq 0\}$. In other words, they are the observations whose the coefficients, associated with the atom $\mathbf{d}_k$ are non zeros. This forces the atom $\mathbf{d}_k$ to fit only a subset of observations and not the entire data set and, thus, enforces sparsity. Then, we compute the error matrix $\mathbf{E}_{w_k} \in \mathcal{R}^{m_s \times \text{card}(w_k)}$. It corresponds to the estimation error between the observation vector $\{\mathbf{s}_i\}_{i \in w_k}$ that forms the columns of $\mathbf{S}_{w_k} \in \mathcal{R}^{m \times \text{card}(w_k)}$ and the signal estimated for the group of observation w_k without taking into account the kth atom, i.e $\tilde{\mathbf{S}}_{w_k} = \sum_{j \neq k} \mathbf{d}_j \mathbf{c}_j^T(i), i \in w_k$. Finally, we use a non linear approach (the LMA) to estimate the polynomials and scale parameters $a_{k,1..n}$, ζ_k , which enable $\mathbf{d}_k$ to fit the error matrix $\mathbf{E}_{w_k}$. The method, as a whole, is given in Alg. 1.

Convergence: The sparse coding step (Eq. 5bis in Alg. 1) is well known to be convex and FISTA allows to converge to the unique solution specific to the current dictionary D. The dictionary update step, where $\mathbf{a_j}$, ζ_j are estimated using the Levenberg-Marquardt algorithm (LMA), could converge to local minima, depending on the initial solutions. In our experiments, a good convergence has been reached after few iterations with the polynomials parameters initialized as Laguerre polynomials coefficients of order n and l and a fixed $\zeta = 700$ [4].

3.2 Reconstruction

The purpose of section 3.1 was to learn the dictionary $\mathbf{D}$. Now, using the estimated $\mathbf{D}$, we are able to model any diffusion signal s , i.e. not in the training data set used to learn D, by solving for c and with FISTA [6], the convex problem $\arg\min_{\mathbf{c}}\{\|\mathbf{s} - \mathbf{Dc}\|_2^2\}$ subject to $\|\mathbf{c}\|_1 \leq \epsilon$, where ϵ is a small number.

4 Results

Synthetic Data from a Multi-tensor Model. The attenuation signal is described, by F fibres, as $E(q\mathbf{u}) = \sum_{f=1}^F p_f \exp(-4\pi^2\tau q^2 \mathbf{u}^T D_f \mathbf{u})$ where a fibre f is defined by a tensor matrix D_f and weight p_f. q denotes the norm of the effective gradient and $\mathbf{u}$ is a unitary vector in Cartesian coordinate.

Firstly, we learn our dictionary with a training data composed of $M = 5000$ synthetic signals, evaluated on $ms = 1000$ q-space samples spread between $b_{min} = 0$ and $b_{max} = 10000$ s/mm^2. For each diffusion signal generated we randomly vary the number of fibers (between 1 and 2 fibers), the fractional anisotropy related to a fiber (between 0.75 and 0.90) and the crossing angle between these fibers (between 30° and 90°). The maximal angular and radial order of the dictionary are respectively set to $L = 8$ and $N = 5$, which gives 270 atoms. The algorithm 1 converges in 9 iterations in about 20 minutes with Python and a CPU at 2.8 GHz.

Secondly, we proceed to the signal reconstruction (see sec. 3.2) using our previous learned dictionary. To evaluate the reconstruction, we compute the normalized mean square error (NMSE $=\|E - \tilde{E}\|_2^2/\|E\|_2^2$) between the original observation signal E and the estimated signal $\tilde{E}$. We consider three cases : one fiber, two fibers crossing at 90° and two fibers crossing at 60°, and generate them using the multi-tensor model. We take 50 samples, a clinically acceptable number of acquisitions, spread on three shells of b-values 500, 1500, 2500 s/mm^2 along a spherical uniform distribution, and add Rician noise with $SNR = 10, 20, 30$. We compare our results while replacing the learned dictionary by the SPF and SHORE bases [2,8] in the reconstruction , because these bases are known to sparsely represent the diffusion signal. We average the results on 1000 trials. To perform a fair comparison, for each trial we try several regularization parameters in the reconstruction (ϵ in sec. 3.2) and keep the lowest NMSE. The averaged NMSEs are shown in table 1. We also add the averaged number of non zero coefficients after reconstruction. It indicates the sparsity of the dictionary/basis used in the reconstruction.

Overall, we obtain a higher sparsity (lower number of atoms) and a higher accuracy (lower NMSE) using the LD. Moreover, for $SNR = 10$ the SPF basis is clearly not appropriate in an ℓ_1 minimization reconstruction. The corresponding NMSE, between the reconstruction using the learned dictionary and the SPF/SHORE basis, gets closer while increasing the SNR. However the number of atoms still remains higher with the SPF and SHORE bases. These results are not surprising while taking synthetic signals as training data set. In the next section, we prove the effectiveness of our method on a phantom data set.

Table 1. NMSE between the estimated signal and the ground truth signal for three different SNR. The reconstruction is based on the Learned Dictionary (top, LD), the SHORE basis (second line) and the SPF basis (third line).

		SNR=10		SNR=20		SNR=30	
		NMSE	atoms number	NMSE	atoms number	NMSE	atoms number
	one fiber	0.019421	11.50	0.007712	16.18	0.004757	18.32
LD	60°- cross. fib.	0.017969	9.12	0.007079	14.04	0.004072	16.16
	90°- cross. fib.	0.015642	6.45	0.006061	9.49	0.003629	12.07
	one fiber	0.026667	16.21	0.009804	22.05	0.005246	25.90
SHORE	60°- cross. fib.	0.023187	13.31	0.009119	19.13	0.004920	24.88
	90°- cross. fib.	0.021361	12.45	0.008370	18.03	0.004569	20.64
	one fiber	0.032988	19.43	0.013062	23.62	0.005901	31.53
SPF	60°- cross. fib.	0.031719	18.79	0.012131	21.34	0.005629	28.69
	90°- cross. fib.	0.026818	14.05	0.011317	16.72	0.005273	24.19
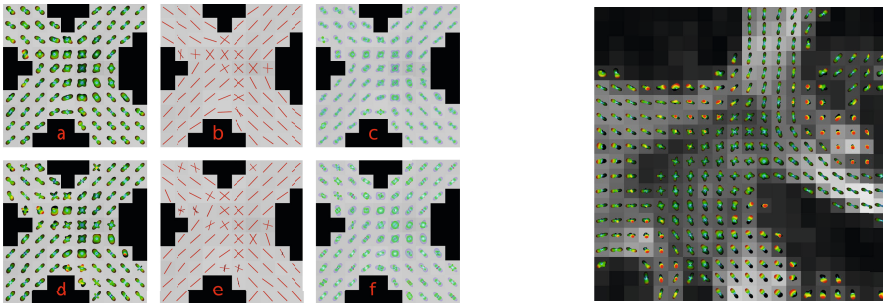
Fig. 1. Left : a, d : ODF estimations. b,e : ODF maxima. c, f : EAP estimations at radii 5 μm (red), 10 μm (green) and 15 μm (blue). Top : Our method. Bottom : SHORE estimation. **Right** : ODF from a coronal slice of a human brain.

Phantom Data. We perform our experiments on a phantom data used in a fiber cup contest in MICCAI 2009 [7]. The data were acquired for three different b-values b=650/1500/2000 s/mm^2, 64 orientations at each b-value, and an imaging matrix of 64x64x3. We use two slices as training data set. The test data set is the third slice. The angular and radial order of the dictionary are respectively $L = 8$ and $N = 5$.

For the reconstruction we take 50 samples following an uniform spherical law. From the estimated signals, we first present in Fig. 4 (left) : a) the ODFs computed via Eq. 4 and d) the ODF computed via the SHORE framework [8]. However, because it is quite difficult to directly give an appreciation on these figures, we compute the local maxima of the ODFs (see Fig. 4.b and e) (left)). The maxima show that the ODFs, based on our method, catch significant angular information, whereas the ODFs based on the SHORE framework do not model the angular information as precisely.

Furthermore, we present in Fig. 4.c and f (left) the EAP computed at three different radii 5 μm (red), 10 μm (green) and 15 μm (blue) respectively using the closed form at Eq. 3 and the SHORE framework [8]. It adds a new dimension to the ODF feature because both radial and angular information are caught. Again the SHORE estimation appears more noisy than the EAP estimated with our method. The EAP fully describes the diffusion process. However, few applications using this feature exist because of the large number of measurements usually required. With our method, we are able to get a continuous approximation of the EAP and ODF with a clinically acceptable number of measurements (50 samples). It is worthwhile to note that since our method requires less atoms to estimate than SHORE and SPF, we could also reconstruct the EAP and ODF with much less samples while being less sensitive to noise than SHORE and SPF.

Real Data. We validate our method on a real data from a human brain. The data were acquired for three different b-values b=500/1000/2000 s/mm^2, 60 orientations at each b-value,and an imaging matrix of 93x116x93. We add a part of the data set in learning process and reconstruct the signal from a coronal

slice with 50 samples following an uniform spherical law. Fig. 4 (right) shows the ODF estimated via our closed form in eq. 4. These ODFs correctly show the represented crossing region.

5   Conclusions

In this work, we proposed an original and efficient approach to exploit the ability of Compressed Sensing (CS) to recover dMRI signals from limited number of samples. Our approach allows to learn a parametric dictionary characterized by a set of polynomial and scale parameters well adapted to sparsely and continuously model the diffusion signal as well as to reconstruct in closed form two of its important features : the EAP and the ODF. We showed that our framework outperforms the SPF and SHORE framework in both ODF and EAP estimations. Other diffusion features such as the probability to return to zero, the mean square displacement, and high order moments features can also be easily derived from our framework.

References

1. Aharon, M., Elad, M., Bruckstein, A.: K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. IEEE Transactions on Signal Processing 54(11), 4311–4322 (2006)
2. Assemlal, H., Tschumperl, D., Brun, L.: Efficient and robust computation of PDF features from diffusion MR signal. Medical Image Analysis 13(5), 715–729 (2009)
3. Cands, E., Wakin, M.: An introduction to compressive sampling. IEEE Signal Processing Magazine 25(2), 21–30 (2008)
4. Cheng, J., Jiang, T., Deriche, R.: Theoretical analysis and practical insights on EAP estimation via a unified HARDI framework. In: MICCAI Workshop CDMRI (2011)
5. Descoteaux, M., Angelino, E., Fitzgibbons, S., Deriche, R.: Regularized, fast, and robust analytical q-ball imaging. Mag. Res. in Med. 58(3), 497–510 (2007)
6. Elad, M., Matalon, B., Shtok, J., Zibulevsky, M.: A wide-angle view at iterated shrinkage algorithms. In: SPIE - Wavelet XII, vol. 6701, pp. 26–29 (2007)
7. Fillard, P., Descoteaux, M., Goh, A., Gouttard, S., Jeurissen, B., Malcolm, J., Ramirez-Manzanares, A., Reisert, M., Sakaie, K., Tensaouti, F., Yo, T., Mangin, J.F., Poupon, C.: Quantitative analysis of 10 tractography algorithms on a realistic diffusion MR phantom. Neuroimage 56(1), 220–234 (2011)
8. Ozarslan, E., Koay, C., Shepherd, T., Blackband, S., Basser, P.: Simple harmonic oscillator based reconstruction and estimation for three-dimensional q-space MRI. In: ISMRM, p. 1396 (2009)
9. Rathi, Y., Michailovich, O., Setsompop, K., Bouix, S., Shenton, M.E., Westin, C.-F.: Sparse Multi-Shell Diffusion Imaging. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part II. LNCS, vol. 6892, pp. 58–65. Springer, Heidelberg (2011)
10. Tristán-Vega, A., Westin, C.-F.: Probabilistic ODF Estimation from Reduced HARDI Data with Sparse Regularization. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part II. LNCS, vol. 6892, pp. 182–190. Springer, Heidelberg (2011)

Resolution Enhancement of Diffusion-Weighted Images by Local Fiber Profiling

Pew-Thian Yap and Dinggang Shen

Department of Radiology and Biomedical Research Imaging Center (BRIC)
The University of North Carolina at Chapel Hill, U.S.A.
{ptyap,dgshen}@med.unc.edu

Abstract. Diffusion-weighted imaging (DWI), while giving rich information about brain circuitry, is often limited by insufficient spatial resolution and low signal-to-noise ratio (SNR). This paper describes an algorithm that will increase the resolution of DW images beyond the scan resolution, allowing for a closer investigation of fiber structures and more accurate assessment of brain connectivity. The algorithm is capable of generating a dense vector-valued field, consisting of diffusion data associated with the full set of diffusion-sensitizing gradients. The fundamental premise is that, to best preserve information, interpolation should always be performed along fiber streamlines. To achieve this, at each spatial location, we probe neighboring voxels in various directions to gather diffusion information for data reconstruction. Based on the fiber orientation distribution (FOD), directions that are more likely to be traversed by fibers will be given greater weights during interpolation and vice versa. This ensures that data reconstruction is only contributed by diffusion data coming from fibers that are aligned with a specific direction. This approach respects local fiber structures and prevents blurring resulting from averaging of data from significantly misaligned fibers. Evaluations suggest that this algorithm yields results with significantly less blocking artifacts, greater smoothness in anatomical structures, and markedly improved structural visibility.

1 Introduction

Due to the nature of diffusion magnetic resonance imaging, acquiring images higher than the typical $2\,\text{mm}$ isotropic resolution is extremely difficult without incurring unrealistic scan times and causing very low SNR due to reduced voxel size. Increasing the resolution, however, is not only important for registration, segmentation, and tractography to be performed with greater accuracy, but is also crucial for better visualization of anatomical structures to identify possible neuropathologies. Solutions to achieve higher resolution include employing higher magnetic fields or stronger/faster gradients, dedicated acquisition techniques [4,8], as well as post-processing algorithms [1,6]. In this paper, we will take the last approach and demonstrate that resolution such as $(1\,\text{mm})^3$ can be achieved purely via post-processing techniques.

A decent super-resolution algorithm called Track-Density Imaging (TDI) was recently proposed by Calamante et al. [1]. TDI gains its image contrast by computing the number of fiber streamlines traversing each element of a high-resolution image grid. In order to generate a super-resolution image, whole-brain probabilistic tractography is first performed (generated by randomly seeding a large number of tracks throughout the brain, e.g., > 1,000,000). From these fiber tracks, the total number of tracks present in each grid element is then calculated. While promising, TDI is limited for the following reasons: 1) The results can only be as accurate as the tractography algorithm employed; 2) Contrast can only be extracted from regions that can be reached by the fiber streamlines (i.e., regions with reasonably high diffusion anisotropy); and 3) Only scalar maps, computed from fiber statistics, can be generated.

In this paper, we present a technique that will exploit the continuity information given by local fiber structures to increase the resolution of DWI data beyond the initial scan resolution. Our algorithm will generate a spatially dense vector-valued field consisting of diffusion data associated with the full set of diffusion-sensitizing gradient directions. Very much in the spirit of TDI, our approach gains spatial resolution by using additional information obtained from outside each individual voxel. Our approach, however, has a slightly different aim with the following important distinctions:

1. **Directional Profiling** — Instead of relying on tractography, our approach uses a directional profiling scheme to study the neighborhood of each spatial location and gauge the probability of whether a specific direction is likely to be traversed by fibers. The directional probability distribution is employed to encourage interpolation along the tangential and not orthogonal directions of fiber streamlines. Unlike the conventional trilinear interpolation approach, which has no concern of the directional structure of DWI data, our approach mimics DWI acquisition mechanism more closely by infusing information from different directions to reconstruct the DWI data at each spatial location.

2. **Microstructure-Preserving Smoothing** — Smoothed diffusion data allow fiber pathways to be tracked more reliably and in a continuous manner. Equally important is the preservation of boundaries that defines the spatial extent of individual structures. In contrast to many existing methods that use *inter*-voxel gradient information to constrain smoothing to relatively homogeneous regions, our method uses *intra*-voxel fiber orientation distributions (FODs) to guide smoothing. By constraining interpolation along fiber streamlines, we not only preserve boundaries of the white matter, but also those between fiber tracts within the white matter.

3. **Complete DWI Data** — Our approach generates a vector-valued field of DWI data corresponding to the full set of diffusion-gradients and is not limited to the white matter. Gray matter provides contextual information for the white matter and the availability of gray matter data allows tissue

segmentation based on diffusion data, such as that done in [5], to be performed, providing complementary tissue contrast to tissue segmentation based on structural MRI.

2   Approach

To increase resolution, the image domain is divided using a grid with grid elements that are smaller than the acquired voxel size. The DWI data for each of these grid elements is then reconstructed using the following steps: 1) Directional profiling in a field of FODs [2]; 2) Interpolation of DWI data based on the FOD profile; and 3) Bias correction associated with the Rician distribution nature of the magnitude signal. Each step is detailed in the following sections.

2.1   Local Fiber Profiling

To determine the probability of whether a grid element at spatial location $\mathbf{x}$ is traversed by fibers in direction $\mathbf{v}_k$ ($k = 1, \ldots, M$), we profile the field of FODs $\{p(\mathbf{x}, \mathbf{v})\}$ in the neighborhood of $\mathbf{x}$ along direction $\mathbf{v}_k$ (see Fig. 1). The FOD profile is defined as:

$$\hat{p}(\mathbf{x}, \mathbf{v}_k) = \frac{\sum_{\mathbf{x}_i \in \mathcal{N}(\mathbf{x}, \mathbf{v}_k)} w(\mathbf{x}_i, \mathbf{x}, \mathbf{v}_k) p(\mathbf{x}_i, \mathbf{v}_k)}{\sum_{\mathbf{x}_i \in \mathcal{N}(\mathbf{x}, \mathbf{v}_k)} w(\mathbf{x}_i, \mathbf{x}, \mathbf{v}_k)} \tag{1}$$

where $\mathcal{N}(\mathbf{x}, \mathbf{v}_k)$ is a neighborhood of voxels in the vicinity of a reference line radiating from $\mathbf{x}$ in the direction $\mathbf{v}_k$. The weights are determined by

$$w(\mathbf{x}_i, \mathbf{x}, \mathbf{v}_k) = \exp\left(-\frac{d_{\mathrm{axial}}^2(\mathbf{x}_i, \mathbf{x}, \mathbf{v}_k)}{2\sigma_{\mathrm{axial}}^2}\right) \exp\left(-\frac{d_{\mathrm{radial}}^2(\mathbf{x}_i, \mathbf{x}, \mathbf{v}_k)}{2\sigma_{\mathrm{radial}}^2}\right) \tag{2}$$

where d_{axial} and d_{radial} are respectively the axial distance (length parallel to the reference line) and the radial distance (length perpendicular to the reference line), as illustrated in Fig. 1. Specifically,

$$d_{\mathrm{axial}}(\mathbf{x}_i, \mathbf{x}, \mathbf{v}_k) = (\mathbf{x}_i - \mathbf{x}) \cdot \mathbf{v}_k \tag{3}$$

and

$$d_{\mathrm{radial}}(\mathbf{x}_i, \mathbf{x}, \mathbf{v}_k) = ||\mathbf{x}_i - \mathbf{x} - d_{\mathrm{axial}}(\mathbf{x}_i, \mathbf{x}, \mathbf{v}_k)\mathbf{v}_k||, \tag{4}$$

where $||\cdot||$ is the Euclidean norm. Parameters σ_{axial} and σ_{radial} control the falloff of the weight with respect to the axial and radial distances, respectively. The resulting FOD profile $\hat{p}(\mathbf{x}, \mathbf{v}_k)$ is a directional function that allows for anisotropic interpolation of neighboring information to reconstruct the DWI data of the grid element at location $\mathbf{x}$.
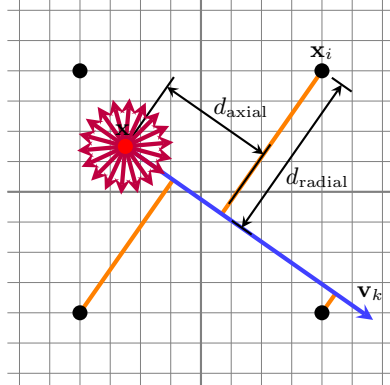
Fig. 1. Directional Profiling. The space resided by the voxels (black dots) are divided by a grid with resolution finer than the scan resolution. For each grid element of interest (red dot), we probe its neighborhood in directions $\mathbf{v}_k$, $k = 1, \ldots, M$ (red arrows) to obtain a FOD profile $\hat{p}(\mathbf{x}, \mathbf{v}_k)$, which is essentially a directional function informing us of the certainty of interpolating in a certain direction. The location $\mathbf{x}$ and direction $\mathbf{v}_k$ defines a reference line. Voxels closer to the reference line (blue), measured with $d_{\mathrm{axial}}(\mathbf{x}_i, \mathbf{x}, \mathbf{v}_k)$ and $d_{\mathrm{radial}}(\mathbf{x}_i, \mathbf{x}, \mathbf{v}_k)$, will be given greater weights for interpolation.

2.2 Fiber-Sensitive Interpolation with Rician-Bias Correction

To reconstruct the DWI data at location $\mathbf{x}$, interpolation is performed with the help of the FOD profile:

$$\hat{S}(\mathbf{x}, \mathbf{g}_l) = \left[\frac{\sum_k \hat{p}(\mathbf{x}, \mathbf{v}_k) R(\mathbf{x}, \mathbf{v}_k, \mathbf{g}_l)}{\sum_k \hat{p}(\mathbf{x}, \mathbf{v}_k)} - 2\sigma_{\mathrm{rician}}^2 \right]_+^{\frac{1}{2}} \tag{5}$$

where

$$[z]_+ = \begin{cases} z, & z > 0; \\ 0, & \text{otherwise,} \end{cases} \tag{6}$$

and

$$R(\mathbf{x}, \mathbf{v}_k, \mathbf{g}_l) = \frac{\sum_{\mathbf{x}_i \in \mathcal{N}(\mathbf{x}, \mathbf{v}_k)} w(\mathbf{x}_i, \mathbf{x}, \mathbf{v}_k) S^2(\mathbf{x}_i, \mathbf{g}_l)}{\sum_{\mathbf{x}_i \in \mathcal{N}(\mathbf{x}, \mathbf{v}_k)} w(\mathbf{x}_i, \mathbf{x}, \mathbf{v}_k)}. \tag{7}$$

To avoid blurring of the reconstructed data by data in directions that have low probability of being traversed by fibers, we consider in (5) only directions with $\hat{p}(\mathbf{x}, \mathbf{v}_k) > \langle \hat{p}(\mathbf{x}, \mathbf{v}_k) \rangle_{\mathbf{v}_k}$. Note that the squared signal values are used here so that the statistical bias $2\sigma_{\mathrm{rician}}^2$ can be removed for unbiased estimation. This is derived from the fact the second order moment of a Rician distributed quantity is given as $E(\mathbf{S}^2) = S_{\mathrm{true}}^2 + 2\sigma_{\mathrm{rician}}^2$, where S_{true} is the true signal value. The noise variance $\sigma_{\mathrm{rician}}^2$ associated with the Rician distribution [7] can be estimated from the background signal ($S_{\mathrm{true}} = 0$) using $\sigma_{\mathrm{rician}}^2 = \sqrt{\langle S_{\mathrm{background}}^2 \rangle / 2}$.

3   Experimental Results

We evaluated the effectiveness of the proposed method using a set of *in vivo* DWI data. For all cases, we set $\sigma_{\text{axial}} = 2$ and $\sigma_{\text{radial}} = 1$. Directional profiling was performed in 642 directions, which were generated by subdividing the faces of an icosahedron 3 times.

3.1   Dataset

Diffusion-weighted images for 4 adult subjects were acquired using a Siemens 3T TIM Trio MR Scanner with an EPI sequence. Diffusion gradients were applied in 120 non-collinear directions with diffusion weighting $b = 2000\,\text{s/mm}^2$, flip angle $= 90°$, repetition time (TR) $= 12,400\,\text{ms}$, and echo time (TE) $= 116\,\text{ms}$. The imaging matrix was 128×128 with a rectangular FOV of $256 \times 256\,\text{mm}^2$. The slice thickness was $2\,\text{mm}$. T_1-weighted structural images were also acquired as anatomical references.

3.2   Computation of the FOD Field

The FOD field was computed by fitting mixtures of tensors to the DWI data. An over-complete set of tensors were used and the fitting problem was solved using an L_1-constrained sparse representation framework [2]. We utilized an active-set-based algorithm that was modified from the feature-sign algorithm presented in [3] to incorporate a non-negative constraint. This approach to FOD estimation is known to have great robustness to noise.

3.3   Methods for Comparison

We up-sampled each DW image from its original resolution of $(2\,\text{mm})^3$ to $(1\,\text{mm})^3$. The results given by the standard trilinear interpolation, performed both on the anisotropy map as well as the vector-valued DWI data, were shown for comparison. The generalized anisotropy at each voxel location was computed as the standard deviation to RMS ratio of the signal values.

3.4   Qualitative Evaluation

A set of representative results are shown in Fig. 2 together with close-up views in Fig. 3. It is clear from the results that the proposed method yields results that exhibit significantly less blocking artifacts, greater smoothness in anatomical structures, and markedly improved structural visibility. Note that interpolation of DWI benefits quite significantly from interpolation schemes that consider the directionality of the data. From the figure we can see that when interpolation is performed on the scalar anisotropy image, the resulting image is blurred with not much structural details. Details are however much more readily visible when interpolation is performed on the vector-valued DWI data, i.e., by independently
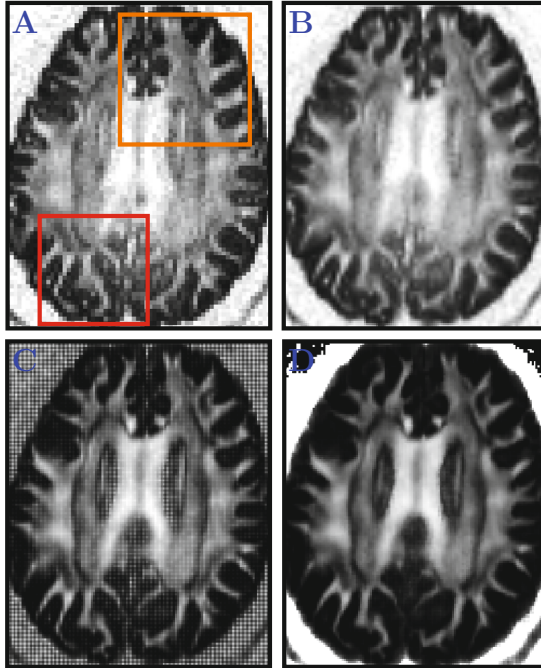
Fig. 2. DWI Resolution Enhancement. DW images were up-sampled to an $(1mm)^3$ resolution from the original $(2mm)^3$ resolution. (A) The anisotropy image of the original data; (B) The linearly up-sampled anisotropy image; (C) The anisotropy image of the linearly up-sampled DWI data; and (D) The anisotropy image given by the proposed method. Close-up views of the regions marked in orange and red are shown in Fig. 3.

interpolating each element of the data vector across space. This is largely due to the fact that by doing so we are essentially computing a convex combination of neighboring diffusion data that is associated with each particular gradient direction. While clearly giving more structural details, vector-based linear interpolation results in significant blockiness due to the Cartesian nature of the interpolation. All these problems are effectively dealt with using the proposed method, as is evident from the results.

3.5   Quantitative Evaluation

For quantitative evaluation, we computed the local variances of the generalized anisotropy values in the white matter (WM), gray matter (GM) and cerebrospinal fluid (CSF). For this purpose, we segmented the T_1-weighted image associated with each of the 4 subjects into WM, GM, and CSF. These segmented images were then used as masks to compute the average anisotropy variances for regions corresponding to WM, GM, and CSF. The variability of the anisotropy values gives us an indication on how smooth the interpolated image is. The
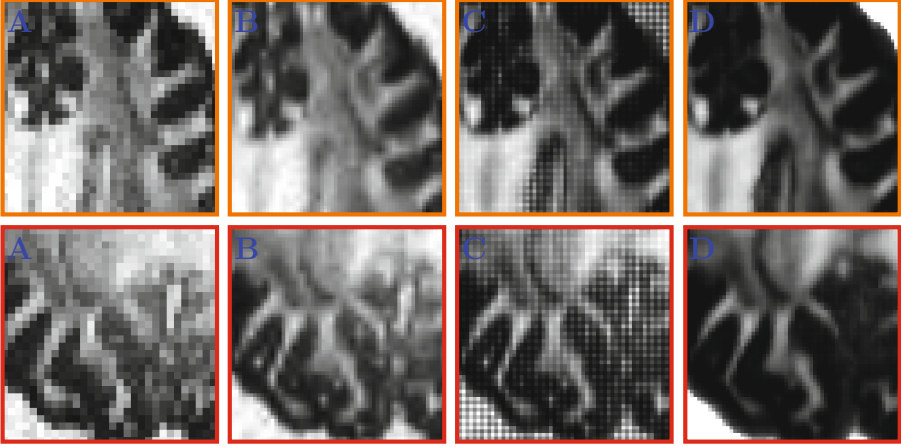
Fig. 3. Closeup Views. Regional closeup views of Fig. 2.
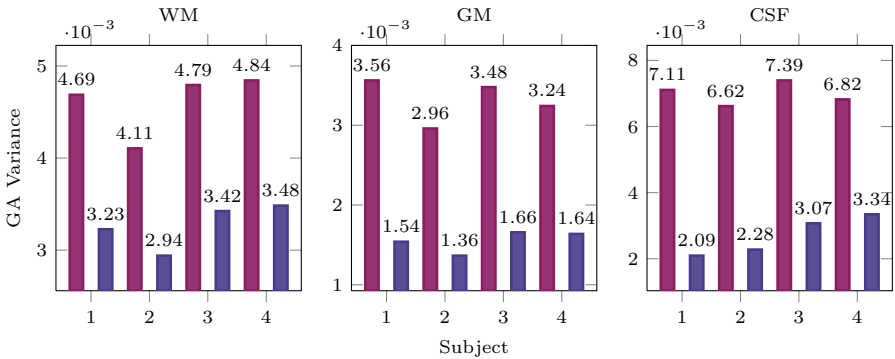
Fig. 4. Average Regional Variance of Generalized Anisotropy. The anisotropy variance at each voxel location is computed using a $3 \times 3 \times 3$ neighborhood. The bars show the average values of the anisotropy variance. Only results for vector-based linear interpolation (■) and the proposed method (■) are shown since these two methods yield better structural clarity.

results, shown in Fig. 4, indicate that the proposed method, while giving markedly improved structural details as shown in the previous section, also yields greater structural smoothness. Such smoothness is important for applications such as tractography to ensure that fiber pathways can be traced more reliably with greater continuity. The blocking artifacts caused by vector-based linear interpolation will cause abrupt changes in local fiber orientations, the effect of which on tractography can be disastrous.

4   Conclusion

We have presented a technique that increases the image resolution of DWI data beyond the acquired resolution, producing results that are free from blocking artifacts and imaging noise while at the same time show excellent structural clarity. Our approach produces spatially dense vector-valued data that are available for all diffusion gradient directions as well as all tissue types. This important distinction essentially implies that, with the results generated by our approach, the myriad of existing diffusion models can be fitted for multifaceted analysis. In addition, our work also prompts the need of a closer examination on the effect of interpolation artifacts on the performance of various DWI algorithms, such as tractography. Future work will be directed at utilizing locally non-linear interpolation strategies to elucidate even finer structural details and to resolve complex fiber architectures such as configurations associated with branching and kissing.

References

1. Calamante, F., Tournier, J., Jackson, G., Connelly, A.: Track-density imaging (TDI): Super-resolution white matter imaging using whole-brain track-density mapping. Neuroimage 53(4), 1233–1243 (2010)
2. Jian, B., Vemuri, B.C.: A unified computational framework for deconvolution to reconstruct multiple fibers from diffusion weighted MRI. IEEE Transactions on Medical Imaging 26(11), 1464–1471 (2007)
3. Lee, H., Battle, A., Raina, R., Ng, A.Y.: Efficient sparse coding algorithms. In: NIPS, pp. 801–808 (2007)
4. Liu, C., Bammer, R., Kim, D.H., Moseley, M.E.: Self-navigated interleaved spiral (SNAILS): Application to high-resolution diffusion tensor imaging. Magnetic Resonance in Medicine 52(6), 1388–1396 (2004)
5. Liu, T., Li, H., Wong, K., Tarok, A., Guo, L., Wong, S.T.: Brain tissue segmentation based on DTI data. NeuroImage 38(1), 114–123 (2007)
6. Nedjati-Gilani, S., Alexander, D., Parker, G.: Regularized super-resolution for diffusion MRI. In: IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pp. 875–878 (2008)
7. Nowak, R.: Wavelet-based Rician noise removal for magnetic resonance imaging. IEEE Transactions on Image Processing 8(10), 1408–1419 (1999)
8. Scherrer, B., Gholipour, A., Warfield, S.K.: Super-Resolution in Diffusion-Weighted Imaging. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part II. LNCS, vol. 6892, pp. 124–132. Springer, Heidelberg (2011)

Geodesic Shape-Based Averaging

M. Jorge Cardoso[1], Gavin Winston[2], Marc Modat[1], Shiva Keihaninejad[3],
John Duncan[2], and Sebastien Ourselin[1,3]

[1] Centre for Medical Image Computing (CMIC), UCL, UK
[2] Epilepsy Society MRI Unit, Dep. of Clinical and Experimental Epilepsy, UCL, UK
[3] Dementia Research Centre (DRC), UCL, UK

Abstract. A new method for the geometrical averaging of labels or
landmarks is presented. This method expands the shape-based averaging
[1] framework from an Euclidean to a geodesic based distance, incorpo-
rating a spatially varying similarity term as time cost. This framework
has unique geometrical properties, making it ideal for propagating very
small structures following rigorous labelling protocols. The method is
used to automate the seeding and way-pointing of optic radiation trac-
tography in DTI imaging. The propagated seeds and waypoints follow a
strict clinical protocol by being geometrically constrained to one single
slice and by guaranteeing spatial contiguity. The proposed method not
only reduces the fragmentation of the propagated areas but also signif-
icantly increases the seed positioning accuracy and subsequent tractog-
raphy results when compared to state-of-the-art label fusion techniques.

1 Introduction

Diffusion magnetic resonance imaging (diffusion MRI) is an imaging technique
that provides insights about the pattern of diffusion of water molecules in the
brain, frequently represented by a tensor. Tractography algorithms have been
proposed to characterise and delimit white matter fibre bundles [2]. However,
diffusion imaging and subsequently tractography techniques are prone to imag-
ing artefacts and algorithmic limitations [3]. Due to these limitations, the most
commonly used technique for tractography initialisation is still manual localisa-
tion of tractography seeds, waypoints and exclusion zones. Nevertheless, diffusion
based image analysis techniques are increasingly used for their ability to charac-
terise white matter connectivity and microstructure [4], possibly leading to the
development of biomarkers for neurodegenerative disease progression[5].

Ideally, one would like to reduce both human interaction time and inter-
rater variability by standardising the tractography procedure using automated
methods. A basic step which is common to many recently proposed automated
seeding and way-pointing procedures is multi-atlas based information propaga-
tion through inter-subject alignment of multiple image volumes. This alignment
involves either a single [6] or multiple reference anatomies [7] and fusion of can-
didate segmentations using a multitude of methods normally adapted from com-
puter vision. For example, Suarez *et al.* [7] proposes to use non-rigid registration

based multi-atlas propagation combined with the STAPLE (Simultaneous Truth and Performance Level Estimation) fusion framework for automated seed propagation. Note that, contrarily to T1 anatomical label fusion frameworks, none of these automated methods have explored the use of image similarity metrics to improve the fusion accuracy. Furthermore, the label fusion methods used in previous works do not permit geometrically restrictive protocols where only a very limited number of voxels and/or slices are allowed.

In this work we present an extension to Shape-based Averaging (SBA). The Euclidean distance used in SBA is replaced by a geodesic distance in order to integrate a image similarity term in the label fusion algorithm. The geodesic SBA enables the robust propagation of sparse and non-overlaping labels and allows the inclusion of geometrical and volumetric constrains. To the best of our knowledge, this paper presents the first framework that incorporates both an image similarity metric and shape-based label fusion.

2   Methods

This section first introduces the classical SBA framework and its unique geometrical properties, followed by a reformulation of the framework using a geodesic distance metric. Finally, the similarity based geodesic time cost function is described within the geodesic SBA framework.

2.1   SBA and Its Geometrical Properties

The intuitive concept behind the original Shape-based Averaging method [1] is that if one wants to fuse a set of candidate labels from a number of classifiers, the fused label would be the one which has the the smallest Euclidean distance to the boundaries of all the candidate labels, i.e. the one with the mean shape.

For K input images over R^m and $k = 1, ..., K$, let $s_k(i)$ be the label value at location i. Each label is a number in the set $\Lambda = \{0, 1, ..., L - 1\}$, where zero without loss of generality represents the image background. Now let a distance maps $d_{k,l}(i)$ be the signed Euclidean distance transform (EDT) from the pixel i in image k to label l. The signed EDT for each label is computed as the distance between pixel i and the edge of label l, with its value being positive if pixel i in on the inside of the edge, and negative if i is on the outside. Formally, $d_{k,l}(i)$ is negative if $s_k(i) = l$ and $d_{k,l}(i)$ is positive if $s_k(i) \neq l$. It is important to note that $d_{k,l}$ is a convex function.

Then, for each label one can then calculate the mean Euclidean distance from pixel i to the edge of all candidate labels in all K images, as

$$D_l(i) = \frac{1}{K} \sum_{k=1}^{K} d_{k,l}(i) \tag{1}$$

These Euclidean distance transforms from different labels are then fused by minimizing the mean distance over all labels, $S(i) = \underset{l \in \Lambda}{\arg\min} \, D_l(i)$. A key feature
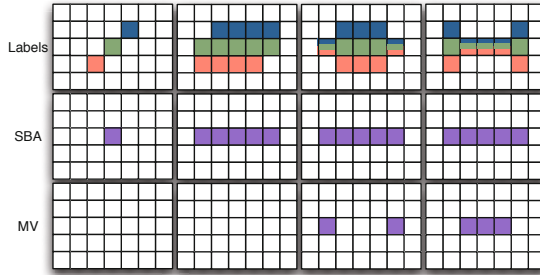
Fig. 1. Top) Four examples with three labels to fuse (in blue, red and green). Overlapping labels are represented by a partial pixels. Middle) Results from the SBA algorithm in purple. Note that the fused label is continuous even if the labels do not overlap. Bottom) Results from the Majority Voting algorithm in purple. Majority voting only places labels in overlaping areas, leading to empty or non-connected results.

to note is that D_l is convex as it is the sum of a series of convex functions $d_{k,l}$. Due to this particularity, SBA is known to produce smooth fusions with reduced structural fragmentations [1]. Also, it is possible to obtain a value of distance from the shape-based average boundary at every pixel and for every label. The combination of the convexity of the function and the fact that one can sample this distance at any point provides unique geometrical properties to the SBA framework. It allows the fusion of labels that do not overlap and labels with completely different shapes. A graphical depiction is show in Fig. 1. These geometrical properties are not present in majority voting, weighted voting and STAPLE label fusion techniques.

2.2   Geodesic Distance Transform

A natural extension of the above described algorithm is the introduction of classifier performance weights. If the proposed algorithm is seen in the perspective of multi-label propagation and fusion, where a series of labels from a template database are propagated to a new unseen image using image registration, then each propagated label should have a different fusion weight depending on how similar the unseen image is to the registered image after transformation. The introduction of classifier performance can help the overall label propagation performance by giving insights about registration accuracy and morphological similarity between images.

In this perspective, the Eq. 1 can then be extended to

$$D_l(i) = \frac{\sum_{k=1}^{K} W_k(i)d_{k,l}(i)}{\sum_{k=1}^{K} W_k(i)} \tag{2}$$

where $W_k(i)$ represent the similarity between the propagated template images k and the current image at position i. If this similarity $W_k(i)$ is spatially varying, one cannot guarantee that $W_k(i)d_{k,l}(i)$ will be monotonically increasing. This
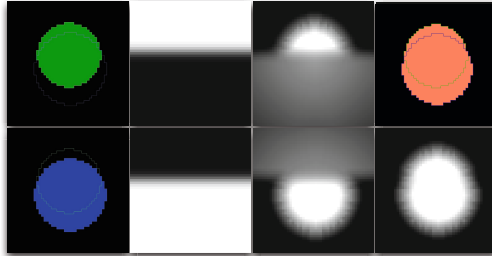
Fig. 2. a) Two labels in blue and green; b) A similarity function $W_k(i)$ for each label with white being similar and black being dissimilar; d) The resultant geodesic distance transforms $d_{k,l}(i)$; d) The final set $S(i)$ on the top and the distance $D_l(i)$ at the bottom. All distances are thresholded between -4 and 4 for the purpose of visualisation.

would only occur if $W_k(i)$ is constant for every i, where $D_l(i)$ is reduced to a weighted sum of Euclidean distance transforms.

In order to maintain the monotonicity of $d_{k,l}(i)$ while introducing a spatially variant similarity term, one can replace the Euclidean distance transform by a Eikonal equation based geodesic distance transform. The Eikonal equation is of the form $\mid \nabla d_{k,l}(i) \mid = W_k(i)$, with $d_{k,l}(i)|_l = 0$ and $W_k(i) > 0$ and with ∇ representing the gradient and $|.|$ representing the Euclidean norm. Physically, $d_{k,l}(i)$ is the shortest time needed to travel from the boundary of label l to i with $W_k(i)$ being the time cost at i. This distance can be solved using the fast marching method [8]. As $d_{k,l}(i)$ is the result of a first order PDE with $W_k(i) > 0$, $d_{k,l}(i)$ is guaranteed to be monotonically increasing. An example of this distance can be seen in Fig. 2.

2.3   Label Propagation and Similarity Metric

Without loss of generality, the proposed method will be used to automate the seeding and way-pointing of optic radiation tractography. Starting from a database of 40 datasets containing both T1 and DTI MRI imaging modalities, each T1 image was rigidly aligned to the fractional anisotropy (FA) of the DTI image using a block matching approach [9]. An expert human rater manually placed the seeds and waypoints on the DTI image, following the protocol defined in [10].

Assuming a new unseen image, all the multimodal datasets in the database can be propagated to this new dataset by preforming an affine registration, using the same block matching approach, followed by a non-rigid alignment step. To make use of the multimodal nature of the available data for the non-rigid registration step, a multi-modal fast free-form registration algorithm [11] was used. As in Daga *et al.* [12], the multimodal data used for the registration step was compromised of the T1 image and the FA image. The resulting transformations were then used to propagate the manually placed seeds and waypoints from each

atlas to the new image under study. These manual labels were resampled using a nearest-neighbour interpolation in order to maintain their binary nature.

The similarity metric used in this work is based on the local sum of squared differences (LSSD) within a gaussian kernel with standard deviation of 5mm. This similarity term is necessary to characterise both the local differences in tissue appearance due to pathology, registration errors and morphological variability between subjects that could not be captured by the registration. The similarity is calculated independently between the propagated T1 and FA images from the atlas and the unseen T1 and FA images, respectively. This results in a FA derived metric ($LSSD_{FA}$) and an T1 derived metric ($LSSD_{T1}$). The LSSD can then be combined into a common similarity metric by using an inverse exponential function. Thus, $W_k(i)$ will be defined as $W_k(i) = e^{(-LSSD_{FA}-LSSD_{T1})}$, where higher values of $W_k(i)$ means higher similarity.

2.4   The Clinical Protocol

The seeding and waypoint placement protocol is described in [10]. In short, the seed mask consists of a series of voxels antero-lateral to the lateral geniculate nucleus at the base of Meyers loop, positioned in the white matter in one single coronal slice. These voxels should be contiguous in order to ensure that the entire coronal cross-section of Meyers loop was encompassed. The volumes of seed masks were standardised to 15 voxels for all subjects in order to reduce inter-subject tractography variability. In addition to the seed, a waypoint was placed in the lateral wall of the occipital horn of the lateral ventricle. This waypoint was also restricted to one single coronal slice.

From this clinical protocol, three main constrains are necessary: both the seeds and waypoints should be continuous and restricted to one single coronal slice and the number of voxels in the seed is standardised to $N_l = 15$. Due to the geometrical nature of the proposed method, these constrains are easily integrated into the current framework by constraining the space of solutions of $S(i)$. Because we are not interested in the background label, instead of finding what is the label that minimises the sum of the geodesic distances, one can find what is the coronal slice with a set $S(i)$ of N_l voxels that minimises the sum of the geodesic distances. As the geodesic $d_{k,l}(i)$ is not necessarily convex, contiguity of $S(i)$ is not mathematically guaranteed. However, the monotonically increasing nature of $d_{k,l}(i)$ will result in a smooth solution.

3   Validation

The validation of the proposed methodology will first quantify the error in automated placement of seeds and waypoints when compared to the manually placed ones using a leave-one-out cross validation approach. The propagated seeds and waypoints will also be characterised in terms of volume and contiguity. Secondly, because small errors in seed positioning can result in drastically different tractography results, the probabilistic overlap between manual and automatically

Table 1. Automated seed and waypoint accuracy results using MV, STAPLE, SBA and the proposed method. Column 'Manual' contains the gold standard statistics. Distance is not available for both MV and SBA because the fused output was empty.

Metric		Seed Fusion					Waypoint Fusion				
		MV	STAPLE	SBA	GSBA	Manual	MV	STAPLE	SBA	GSBA	Manual
Distance	Mean	-	1.70	-	1.60	-	-	1.86	-	1.86	-
	Std	-	1.81	-	1.41	-	-	1.12	-	0.97	-
Voxels #	Mean	0	5.8	0	15	15	0	31.89	0	113.15	136.17
	Std	0	2.89	0	0	0	0	18.85	0	17.62	85.6
Mean # Conn.Comp.		0	1.2	0	1	1	0	1.9	0	1	1
Mean Euler		0	3.2	0	2	2	0	7.2	0	2	2

generated tracts is assessed. For all experiments, the geodesic SBA fusion performance is also compared to majority voting (MV), STAPLE [7], Euclidean SBA [1] with the original cost function for $S(i)$.

3.1 Seed and Way-Point Placement Accuracy

This section aims at assessing both accuracy of seed and waypoint placement and their geometrical properties. In order to do so, both seeds and waypoints were propagated as described in Section 2.3 and then fused using the above described methods. Due to the very small size of the seeds and waypoints and because they are limited in thickness to one single coronal slice, overlap measures like the Dice coefficient would not provide insightful information about the placement accuracy. Instead, we use the mean Euclidean distance from all the points in the automated seed and waypoints to the manually placed ones as a measure of accuracy. We also calculate the number of voxels, the number of connected components and the Euler characteristic of each propagated seed. The later describes topology of the propagated seed, where an Euler characteristic of 2 means that the seed is homotopic to a closed disk.

Results are presented in Table 1. Due to the label positioning variability, both MV and SBA methods resulted in an empty set. Note that SBA also fails because the original metric for $S(i)$ was used. Both GSBA and STAPLE obtain similar results regarding positioning accuracy, but the geometrical characteristics of the STAPLE's seeds and waypoints are discrepant when compared to the gold standard. The mean number of voxels, number of connected components and the Euler characteristic for the seeds and waypoints from STAPLE was significantly different ($p < 10^{-4}$) from the gold standard, while no significant differences were found between the GSBA and the gold standard.

3.2 Automated Tractography Validation

As a geometrically accurate positioning of the seeds and waypoints is meaningless if they lead to different tractography results, a further assessment of the similarity between manually generated and automatically generated tracks is also preformed. In order to do so, the Probabilistic Index of Connectivity (PICo)
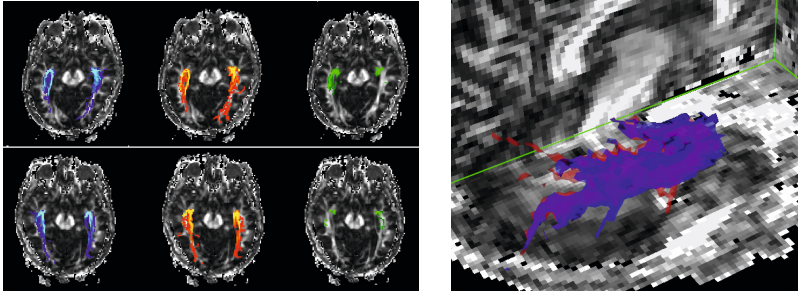
Fig. 3. Left) Optic radiation tractography for two patients. Tractography from the manual seeds, GSBA and STAPLE are in blue, red and green respectively. Right) A 3D recontruction of the optic radiation for the manual and geodesic SBA methods.

algorithm [2] was used for tractography. Voxels in which a single tensor fitted the data poorly were identified using a spherical-harmonic voxel-classification algorithm [13]. In these voxels, a two tensor model was fitted, with the principal diffusion directions of the two diffusion tensors providing estimates of the orientations of the crossing fibres. In all other voxels, a single tensor model was fitted. Tracking from the seed was performed using 50000 Monte Carlo iterations, an angular threshold of 180 (sufficient angular flexibility to allow tracking of Meyer's loop) and a fractional anisotropy threshold of 0.1. Because both MV and SBA method fail to generate usable seeds and waypoints, only STAPLE and the geodesic SBA were used for the comparison.

As the output from PICo is a probabilistic segmentation and because thresholding a probabilistic segmentation introduces too much boundary variability, the probabilistic Dice coefficient [14] was used as a measure of similarity between tracts. Results show that the mean(std) Dice coefficient between the manually seeded tracts and the automated tracts using the geodesic SBA is 0.547(0.138), while with STAPLE method is 0.226(0.114). Using a two-tailed paired t-test for statistical comparison, the geodesic SBA method provides highly statistically significant improvements ($p < 10^{-4}$) in tractography accuracy. Note that this is the overlap of the full tract from the Meyer's loop to the visual cortex.

4   Conclusion

This work presents an geodesic extension the Euclidean Shape-based Averaging (SBA) framework, integrating an image similarity term in the label fusion algorithm. The geodesic SBA enables the robust propagation of sparse and non-overlaping labels and naturally permits the inclusion of geometrical and volumetric constrains into the propagated label sets. Application to DTI tractography shows statistically significant improvements in both seed and waypoint placement accuracy, geometric characteristics and topology when compared to state-of-the-art methodology, leading to improvements in tractography accuracy.

References

1. Rohlfing, T., Maurer, C.R.: Shape-based averaging. IEEE TIP 16(1), 153–161 (2007)
2. Parker, G.J.M., Haroon, H.A., Wheeler-Kingshott, C.A.M.: A framework for a streamline-based probabilistic index of connectivity (PICo) using a structural interpretation of MRI diffusion measurements. J. Magn. Reson. Imaging 18(2) (2003)
3. Bammer, R., Acar, B., Moseley, M.E.: In vivo MR tractography using diffusion imaging. European Journal of Radiology 45(3), 223–234 (2003)
4. Parker, G.J.M., Alexander, D.C.: Probabilistic Monte Carlo Based Mapping of Cerebral Connections Utilising Whole-Brain Crossing Fibre Information. In: Taylor, C.J., Noble, J.A. (eds.) IPMI 2003. LNCS, vol. 2732, pp. 684–695. Springer, Heidelberg (2003)
5. Goldberg-Zimring, D., Mewes, A.U.J., Warfield, S.K.: Diffusion Tensor Magnetic Resonance Imaging in Multiple Sclerosis. Neuroimaging (2005)
6. Zhang, W., Olivi, A., Hertig, S.J., van Zijl, P., Mori, S.: Automated fiber tracking of human brain white matter using diffusion imaging. NeuroImage 42(2) (2008)
7. Suarez, R.O., Commowick, O., Prabhu, S.P., Warfield, S.K.: Automated delineation of white matter fiber tracts with a multiple region-of-interest approach. NeuroImage 59(4), 3690–3700 (2012)
8. Sethian, J.A.: A fast marching level set method for monotonically advancing fronts. PNAS 93(4), 1591–1595 (1996)
9. Ourselin, S., Roche, A., Prima, S., Ayache, N.: Block Matching: A General Framework to Improve Robustness of Rigid Registration of Medical Images (2000)
10. Yogarajah, M., Focke, N.K., Bonelli, S., Cercignani, M., Acheson, J., Parker, G.J.M., Alexander, D.C., McEvoy, A.W., Symms, M.R., Koepp, M.J., Duncan, J.S.: Defining Meyer's loop-temporal lobe resections, visual field deficits and diffusion tensor tractography. Brain 132(6), 1656–1668 (2009)
11. Modat, M., Ridgway, G.R., Taylor, Z.A., Lehmann, M., Barnes, J., Hawkes, D.J., Fox, N.C., Ourselin, S.: Fast free-form deformation using graphics processing units. CMPB 98(3), 278–284 (2010)
12. Daga, P., Winston, G., Modat, M., White, M., Mancini, L., Cardoso, M., Symms, M., Hawkes, D., Duncan, J., Ourselin, S.: Accurate Localisation of Optic Radiation during Neurosurgery in an Interventional MRI Suite. IEEE TMI (December 2011)
13. Alexander, D.C., Barker, G.J., Arridge, S.R.: Detection and modeling of non-Gaussian apparent diffusion coefficient profiles in human brain data. Magn. Reson. Med. 48(2), 331–340 (2002)
14. Crum, W.R., Camara, O., Hill, D.L.G.: Generalized Overlap Measures for Evaluation and Validation in Medical Image Analysis. IEEE TMI 25(11) (2006)

Multi-scale Characterization
of White Matter Tract Geometry

Peter Savadjiev[1,2], Yogesh Rathi[2], Sylvain Bouix[2],
Ragini Verma[3], and Carl-Fredrik Westin[1]

[1] Laboratory for Mathematics in Imaging and [2] Psychiatry Neuroimaging Laboratory
Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA
[3] Section of Biomedical Image Analysis, Dept. of Radiology, University of
Pennsylvania, Philadelphia, PA, USA

Abstract. The geometry of white matter tracts is of increased interest
for a variety of neuroscientific investigations, as it is a feature reflective of
normal neurodevelopment and disease factors that may affect it. In this
paper, we introduce a novel method for computing multi-scale fibre tract
shape and geometry based on the differential geometry of curve sets. By
measuring the variation of a curve's tangent vector at a given point in all
directions orthogonal to the curve, we obtain a 2D "dispersion distribu-
tion function" at that point. That is, we compute a function on the unit
circle which describes fibre dispersion, or fanning, along each direction
on the circle. Our formulation is then easily incorporated into a contin-
uous scale-space framework. We illustrate our method on different fibre
tracts and apply it to a population study on hemispheric lateralization
in healthy controls. We conclude with directions for future work.

1   Introduction

The brain consists of diverse structures, each with a characteristic shape and an
intricate architecture. Their shape varies across the normal population, and is
an important feature thought to reflect genetic and environmental factors that
may contribute to disorders of neurodevelopmental origin or neurodegenerative
diseases (e.g., [1]). In this context, the study of white matter geometry is of
importance to the neuroscience of white matter and disorders that affect it.

A large group of methods for white matter geometry analysis in diffusion MRI
compute the curvature and torsion of individual fibres recovered with a tractog-
raphy algorithm (e.g., [2]). The geometry of *sets* of curves is usually obtained
by mapping individual curves to medial axes/surfaces (e.g. [3]) or an average
representation (e.g., [4,5]). However, this type of mapping may involve heuristic
decisions in the choice of corresponding points and fibre similarity measures. An
elegant alternative was recently introduced on the basis of the currents frame-
work [6], which represents fibre tracts as a smooth vector field and captures
global tract shape while avoiding the need for specific point correspondences [7].

All geometry analysis methods based on fibre tractography are inherently lim-
ited in that tractography does not, in general, produce stable and reproducible

results. Recognizing this limitation, a method for computing white matter geometry indices directly from diffusion imaging data without requiring prior tractography was proposed in [8]. This method, however, is currently only defined for the single tensor model of diffusion, itself with well known limitations.

In the present work, we propose a scale-based white matter geometry analysis method that is situated logically between that of [8], and vector field based methods such as [7]. While our method is based on vector fields derived from tractography and is therefore subject to all associated limitations, its advantage with respect to [8] is that it is not limited by the tensor model of diffusion, and also allows for a more precise characterisation of fibre fanning at different spatial scales, as detailed in Section 1.1. As for the currents method of [7], it is optimized to capture global tract shape and its modes of variation in a population. In contrast, our method computes local geometrical features based on the differential geometry of curve sets, which makes possible subsequent tract-based statistical analysis with methods such as [3]. We note, however, that our method is complementary to the currents framework of [7], and both may be used in conjunction in order to analyse the geometry of a currents vector field.

1.1   Comparison to Tensor-Based Model of Dispersion

Both our method and the method of [8] compute macrostructural white matter geometry. However, they approach this problem in diametrically opposed ways, which makes them complementary to each other. The method of [8] works directly with DTI data within a 3D neighborhood, without any prior knowledge about fibre tracts. This is an advantage, as it allows to avoid tractography and its limitations. This, however, comes at a price: the geometry of all the white matter present in the 3D neighborhood is represented by a single scalar, which may be less informative when distinct fibre populations pass near each other.

In contrast, our new method works with the tangent vector field of tracts obtained by tractography. This reliance on tractography is a limitation, but it also allows the analysis of a specific tract independently of the influence of other nearby tracts. Furthermore, no constraints are imposed on the underlying model of diffusion. Unlike the method of [8], defined only for diffusion tensor fields, our method can be used even with high quality HARDI data. Finally, instead of computing a single scalar to represent fibre dispersion at a point, we compute *directional* dispersion in a "dispersion distribution function", as detailed below.

The two methods therefore exploit the two sides of a basic trade-off: avoiding the uncertainty inherent in tractography [8], vs. exploiting information about tract structure provided by tractography, as proposed here.

2   Geometrical Framework

The work in [9] models white matter fibre geometry by associating an orthonormal frame (E_T, E_N, E_B) with each point along each 3D curve. Here E_T is the curve's tangent vector, and E_N and E_B are the normal and binormal vectors. The

local tract geometry can then be characterized, *with respect to the local frame*, by computing the variation of E_T in the frame's three orthonormal directions.

In this work, we propose to measure the rate of change of the tangent vector E_T not only in the direction of the normal E_N and the binormal E_B, but in *all* directions in the plane orthogonal to E_T. This allows to *avoid* the need to define Frenet frames, which can be unstable. By computing the rate of change of E_T along the entire circle orthogonal to E_T, instead of only 2 directions, we insure against missing "interesting events", and we describe more completely the complex geometry of white matter tracts. The function on a circle thus computed is a 2D "dispersion distribution function" (DDF), i.e. a function on the circle with values proportional to the amount of fibre dispersion in each direction orthogonal to the fibre. This function provides a richer description of dispersion than the scalar measure of [8], and constitutes our main contribution.

3   Approach

3.1   Problem Statement

The goal of our method is the following: given a set $\mathcal{C}$ of curves C_i that represent the output of some streamline tractography algorithm, with a tangent vector E_T defined at each point $\mathbf{p}$ on each curve $C_i \in \mathcal{C}$, we compute the function

$$\Psi(\theta) = \nabla_{\mathbf{v}} E_T. \tag{1}$$

Here $\nabla_{\mathbf{v}} E_T$ represents the covariant derivative of the tangent vector E_T in direction $\mathbf{v}$, with the constraints that $\{\mathbf{v} | \mathbf{v} \in \mathcal{S}^1 \bigcap \mathbf{v} \perp E_T\}$, i.e., $\mathbf{v}$ is a direction on the unit circle centered at $\mathbf{p}$ in the plane orthogonal to E_T, and is denoted by angle θ in this plane. This function is computed at each $\mathbf{p}$ along each C_i, resulting in a DDF at each $\mathbf{p}$ describing the local fibre dispersion pattern relative to the local tangent vector E_T, as illustrated in Fig. 1 (left).

We note that (1) is defined only for continuous vector fields. That is, a line passing though point $\mathbf{p}$ on curve C_i with direction $\mathbf{v}$ must always intersect another curve C_j at point $\mathbf{p}'$ with a tangent vector E_T' defined at $\mathbf{p}'$, otherwise $\nabla_{\mathbf{v}} E_T$ will not exist. Since we work with a vector field $F = \bigcup_{\mathcal{C}} E_T$ of tangent vectors of 1D curves in 3D Euclidean space $\mathbb{R}^3$, F is not continuous, specific point correspondences cannot always be established and (1) is not always defined.

To resolve this issue, we construct a continuous vector field $F^c(x, y, z)$ at each $(x, y, z) \in \mathbb{R}^3$ by averaging F over a neighborhood $\mathcal{N}$ centered at (x, y, z):

$$F^c(x, y, z) = \frac{1}{N} \sum_{(i,j,k) \in \mathcal{N}} F_{(i,j,k)} \tag{2}$$

where $\mathcal{N}$ is some neighborhood of $\mathbb{R}^3$ centered at location (x, y, z), and N is the number of vectors of F that occur within $\mathcal{N}$.

We note that F is discrete because it is obtained from a discrete set of curves, but also because of the discrete nature of diffusion MRI data, which causes
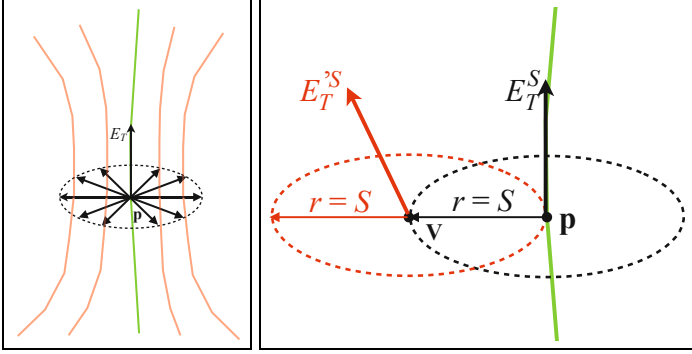
Fig. 1. LEFT: Our method computes at each point **p** a function on the unit circle orthogonal to the tangent vector E_T at **p** which measures the rate of change of E_T in all of these directions, thus computing a 2D "dispersion distribution function". RIGHT: To compute $\nabla_{\mathbf{v}} E_T$ at location **p** at scale S, we define a disk-shaped neighborhood with radius S around **p**, over which vector field $F = \bigcup_{\mathcal{C}} E_T$ is averaged. The same averaging is performed over a neighborhood centered at $\mathbf{p} + S\mathbf{v}$ (orange). $\nabla_{\mathbf{v}} E_T$ is then approximated as the angular difference between the two vectors thus obtained.

each curve C_i not to be continuous, but rather a polyline approximation to a continuous curve. Each C_i is therefore a sequence of small linear segments, with one tangent vector E_T per segment. There is thus a countable number of vectors belonging to F in each neighborhood $\mathcal{N}$, which allows the summation in (2). The spatial location of these vectors is denoted with the (i, j, k) subscript in (2), which are floating point coordinates with sub-voxel resolution.

We also note that *a priori*, the tangent vectors can be presented either as E_T or as $-E_T$. To avoid arbitrary sign changes from one location to the next which can drastically affect the result of (2), we ensure a sign consistency over the entire dataset relative to a global coordinate frame, such as that derived from the principal components of the tracts' spatial distribution.

3.2   Scale Space

Since we seek to characterize the variation of curve orientation in directions orthogonal to the curves, in our implementation we choose the neighborhood $\mathcal{N}$ to be shaped as a *disk* lying in the plane orthogonal to E_T, with a small thickness chosen to be 1 voxel. The radius of this disk is treated as a scale parameter. To compute (1) for a given direction **v** at a given location **p**, we first apply (2) at **p** in order to obtain a value for E_T at scale S, E_T^S. Then, we apply (2) at location $\mathbf{p} + S\mathbf{v}$ (i.e., at a distance S from **p** in direction **v**), in order to obtain a value for the tangent vector $E_T^{'S}$ at location $\mathbf{p} + S\mathbf{v}$. Finally, we approximate $\nabla_{\mathbf{v}} E_T$ (at scale S) as the angular difference between E_T^S and $E_T^{'S}$:

$$\nabla_{\mathbf{v}} E_T \approx \arccos\left(\left\langle E_T^S, E_T^{'S}\right\rangle\right), \tag{3}$$

with $\langle ., . \rangle$ denoting the standard dot product in $\mathbb{R}^3$.
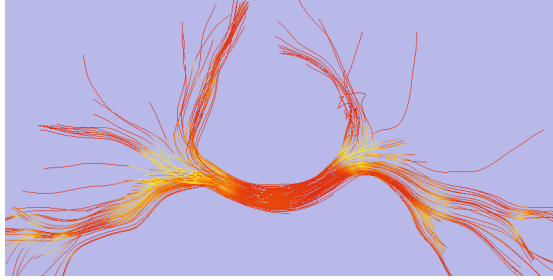
Fig. 2. The total dispersion (TD) visualized for tracts forming part of the corpus callosum. Here and in subsequent figures, yellow indicates high TD values.

These steps are illustrated in Fig. 1 (right). The inclusion of a scale parameter in this manner allows for a simple way to construct a *continuous* scale space.

4   Validation

Our method is a high-level analysis that is applied *after* tractography. We assume the particular tractography method has already been tested on synthetic data or phantoms, and its limitations are known. The focus of the experiments in this paper is the macrostructural geometric analysis of already-traced fibre tracts.

Our method was tested on tracts traced with a filtered tractography algorithm for HARDI data [10]. This method was run on diffusion-weighted imaging data acquired from a volunteer on a GE Signa HDxt 3.0T scanner using an echo planar imaging sequence with a double echo option, an 8 Channel coil and ASSET with a SENSE-factor of 2. The acquisition consisted of 51 directions with b = 900 s/mm^2, and 8 images with b = 0 s/mm^2, with scan parameters TR=17000 ms, TE=78 ms, FOV=24 cm, 144 × 144 encoding steps, 1.7 mm slice thickness. 85 axial slices covering the whole brain were acquired.

As described previously, our method computes 2D DDFs at each point along each curve in the tract set. In order to summarize this large amount of information and present it visually, in the following figures we show fibres where each point is colored by a measure of total dispersion (TD), such that yellow indicates high TD values. We define TD as the average value of $\Psi(\theta)$ (1) at a point.

In Fig. 2, we show the TD measure for a set of fibres passing through the corpus callosum. Here TD was computed with a scale parameter $S = 5$mm. As expected, the measure is highest in regions with highest overall dispersion.

We next explore the effect of varying the scale parameter S on a fibre tract which connects the substantia nigra of the brain stem to the caudate nucleus, a sub-cortical grey matter structure. This tract was selected for illustration purposes because it presents a well-defined fanning structure.

In Fig. 3, we present views of this tract colored by TD, computed at scales S ranging from 1.7mm to 13.6mm. At the smallest scale only very local dispersion features are highlighted. With increasing scale, the larger fanning structure is highlighted more strongly, while small features are progressively lost.

(a) $S = 1.7$mm   (b) $S = 3.4$mm   (c) $S = 5.1$mm   (d) $S = 6.8$mm

(e) $S = 8.5$mm   (f) $S = 10.2$mm   (g) $S = 11.9$mm   (h) $S = 13.6$mm
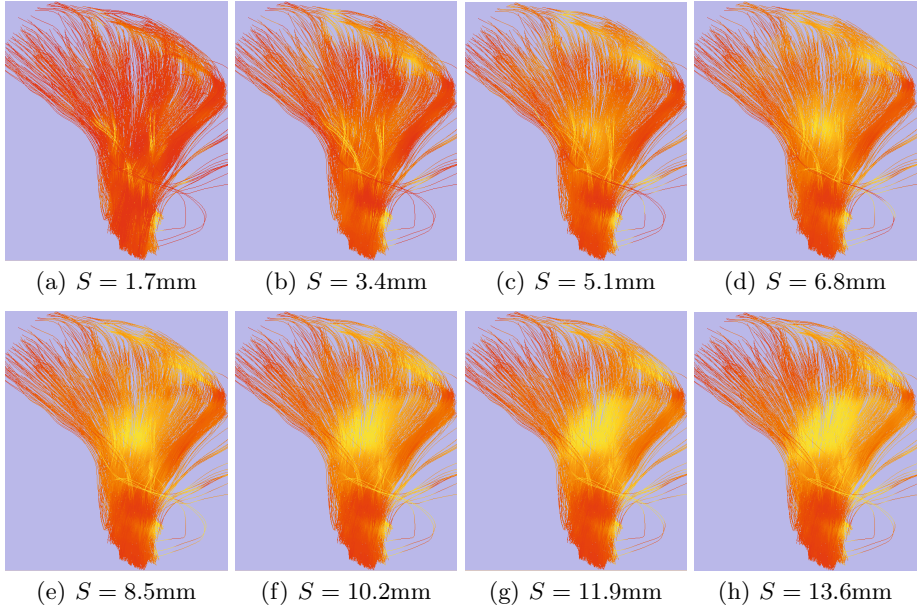
Fig. 3. A fibre tract colored by TD, for a range of values for scale parameter S

5   A Study of Lateralization in Healthy Controls

We illustrate the applicability of the method to population studies via an investigation of lateralization in healthy controls. Diffusion MRI data was acquired from 16 adult healthy male volunteers using the protocol described in Section 4.

Our study was focused on tracts associated to the inferior frontal cortex, as this area has been shown to be lateralized in healthy males, for example in terms of functional connectivity [11], and it has also been implicated in disorders such as autism and schizophrenia (e.g., [11]). We extracted fibre tracts connecting the *pars orbitalis* cortical area (part of the inferior frontal cortex) of each hemisphere, using the filtered HARDI tractography algorithm of [10]. We first computed whole-brain tractography, from which we extracted the interhemispheric tracts connecting both cortical regions as defined by an automated FreeSurfer (http://surfer.nmr.mgh.harvard.edu) parcellation of the cortex. The tracts were then cut within 5 mm from the midsagittal plane. For each subject, we computed the mean TD at scale $S=5$mm over these tracts in each hemisphere. An example of these tracts for one subject is shown in Fig. 4.

The lateralization results are presented in Fig. 5 and they indicate an overall increase of TD in the right hemisphere, with a p-value of 0.030 (two-tailed T test). The male inferior frontal cortex has been previously shown to be right-lateralized in terms of volume and functional connectivity (e.g., [11]). Anatomical differences in white matter geometry are less well known, and our method can be
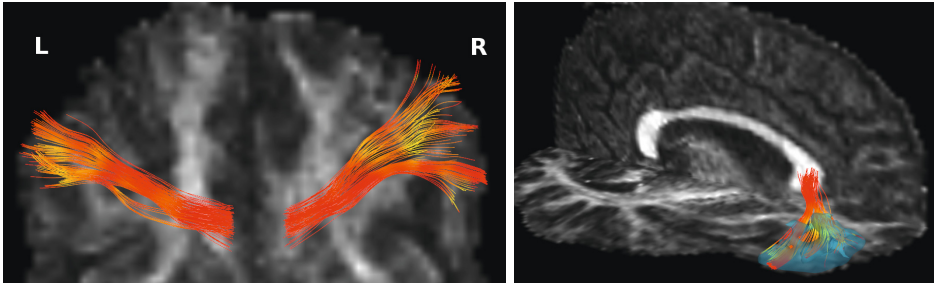
Fig. 4. LEFT: The set of tracts used in our group study shown for one subject. The fibres are colored by TD (computed at scale S=5mm), and are overlaid on an axial slice through the FA volume. RIGHT: The same fibre tract shown with both a sagittal and an axial slice though FA, as well as a semi-transparent model (blue) of the FreeSurfer cortical segmentation of the *pars orbitalis*.
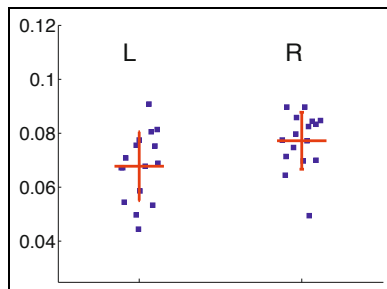
Fig. 5. Comparison of the mean TD value (units of rad/mm) for each individual, shown for each hemisphere. A t-test between the two groups yields a p-value of p=0.030. Horizontal red lines: group mean. The error bars indicate $\pm$ 1 standard deviation. L: left hemisphere. R: right hemisphere.

used in this context. While our current results are preliminary, they do illustrate the applicability of the method to population studies.

6   Discussion and Conclusion

In this paper, we presented a multi-scale approach for computing white matter fibre geometry, based on the local differential geometry of curve sets. The method works with curves traced by fibre tractography algorithms, and has both advantages and weaknesses. The main drawback of this method is its dependence on the quality of the tractography algorithm used to generate the fibres. With this limitation in mind, we reviewed in Section 1.1 several advantages of our method relative to a non-tractography method such as [8]. In addition, we note that fibre tracts provide an explicit correspondence between the white matter and locations on the cortical surface where they originate or terminate. This is very

important for establishing a connection between white matter and grey matter geometry, as the relationship between the two may provide novel insights into a variety of neuroscientific applications, regarding for example brain development or atrophies caused by disease. We will address these questions in future work.

Finally, we note that the TD measure reduces the information in $\Psi(\theta)$ to a directionless scalar. In order to better take advantage of the directional information contained in $\Psi(\theta)$, one may define an inner product between $\Psi(\theta)$ and a vector field, in order to measure dispersion in a specific direction. We will analyse such directional information in future work.

References

1. Toga, A.W., Thompson, P.M.: Mapping brain asymmetry. Nature Rev. Neurosci. 4, 37–48 (2003)
2. Batchelor, P.G., Calamante, F., Tournier, J.D., Atkinson, D., Hill, D.L.G., Connelly, A.: Quantification of the shape of fiber tracts. Magn. Res. in Medicine 55, 894–903 (2006)
3. Yushkevich, P.A., Zhang, H., Simon, T.J., Gee, J.C.: Structure-specific statistical mapping of white matter tracts. NeuroImage 41, 448–461 (2008)
4. Corouge, I., Fletcher, P.T., Joshi, S., Gouttard, S., Gerig, G.: Fiber tract-oriented statistics for quantitative diffusion tensor MRI analysis. Medical Image Analysis 10(5), 786–798 (2006)
5. O'Donnell, L., Westin, C.F.: Automatic tractography segmentation using a high-dimensional white matter atlas. IEEE Trans. Medical Imaging 26(11), 1562–1575 (2007)
6. Vaillant, M., Glaunès, J.: Surface Matching via Currents. In: Christensen, G.E., Sonka, M. (eds.) IPMI 2005. LNCS, vol. 3565, pp. 381–392. Springer, Heidelberg (2005)
7. Durrleman, S., Pennec, X., Trouvé, A., Ayache, N.: Statistical models on sets of curves and surfaces based on currents. Medical Image Analysis 13(5), 793–808 (2009)
8. Savadjiev, P., Kindlmann, G.L., Bouix, S., Shenton, M.E., Westin, C.F.: Local white matter geometry from diffusion tensor gradients. NeuroImage 49, 3175–3186 (2010)
9. Savadjiev, P., Zucker, S.W., Siddiqi, K.: On the differential geometry of 3D flow patterns: Generalized helicoids and diffusion MRI analysis. In: Proc. IEEE Intl. Conf. on Computer Vision, ICCV 2007 (2007)
10. Malcolm, J.G., Shenton, M.E., Rathi, Y.: Filtered multi-tensor tractography. IEEE Trans. on Medical Imaging 29, 1664–1675 (2010)
11. Tomasi, D., Volkow, N.D.: Laterality patterns of brain functional connectivity: Gender effects. Cerebral Cortex 22(6), 1455–1462 (2012)

Optimization of Acquisition Geometry
for Intra-operative Tomographic Imaging

Jakob Vogel[1], Tobias Reichl[1], José Gardiazabal[1],
Nassir Navab[1], and Tobias Lasser[1,2]

[1] Chair for Computer Aided Medical Procedures (CAMP), Technische Universität
München, Germany
[2] Institute for Biomathematics and Biometry, HelmholtzZentrum München, Germany

Abstract. Acquisition geometries for tomographic reconstruction are
usually densely sampled in order to keep the underlying linear system
used in iterative reconstruction as well–posed as possible. While this
objective is easily enforced in imaging systems with gantries, this issue
is more critical for intra–operative setups using freehand–guided data
sensing. This paper investigates an incremental method to monitor the
numerical condition of the system based on the singular value decomposition of the system matrix, and presents an approach to find optimal
detector positions via a randomized optimization scheme. The feasibility
of this approach is demonstrated using simulations of an intra–operative
functional imaging setup and actual robot–controlled phantom experiments.

1   Introduction

Three–dimensional imaging modalities such as X–ray CT, PET or SPECT have
turned out to be indispensable tools for diagnosis in modern medicine. Their
size requirements due to a big imaging gantry, however, prevents wide-spread
use also in intra–operative therapy. In the operating room, the state of the art
of imaging mostly encompasses the use of one– and two–dimensional modalities
like gamma probes or ultrasound, possibly combined with previously acquired
volumetric images.

A possible solution to provide intra–operative volumetric imaging are space–
efficient acquisition setups using small detectors guided by humans (e.g. freehand
ultrasound or freehand SPECT [8]) or robots (e.g. laparoscopic ultrasound with
the Da Vinci robot [6] or C–arm X–ray imaging [2]), and using these measurements for tomographic reconstruction. In this case, it is essential to quickly
generate useful acquisition trajectories such that a minimal number of acquisitions yields the best possible sampling of the region of interest, and thus an
optimal reconstruction for the task at hand.

In this paper we will present an incremental optimization approach based on
the singular value decomposition of the system matrix to compute optimized
acquisition trajectories. The feasibility of this approach is demonstrated using
simulations of an intra–operative functional imaging setup and actual robot–
controlled phantom experiments.

1.1   Terminology

We consider a series expansion approach [4] for iterative reconstruction. In a volume of interest $\Omega \subset \mathbb{R}^3$ the unknown signal $f : \Omega \to \mathbb{R}$ is approximated using a set of n basis functions $b_i : \Omega \to \mathbb{R}$, such that $f \approx \sum_{i=1}^{n} x_i b_i$ with a corresponding coefficient vector $\mathbf{x} = (x_i) \in \mathbb{R}^n$. $\mathbf{x}$ is informally referred to as the *reconstruction* of f. Popular choices for basis functions are for example voxels, Kaiser–Bessel functions, or wavelets.

The set of *measurements* from the detector is denoted $\mathbf{m} = (m_j) \in \mathbb{R}^m$. In order to relate these sensor readings with the image function f, we define physical *measurement models* $\mathcal{M}_j$ that map f to the corresponding value $\mathcal{M}_j(f) = m_j$. In the case of X-ray CT, for example, the Radon transform is a suitable model.

Assuming the models $\mathcal{M}_j$ are linear, the series expansion approach combines the discretization and the physical model:

$$m_j = \mathcal{M}_j(f) \approx \mathcal{M}_j\Big(\sum_i x_i \mathbf{b}_i \Big) = \sum_i x_i \underbrace{\mathcal{M}_j(\mathbf{b}_i)}_{=: \, a_{ji}}.$$

For a specific measurement model $\mathcal{M}_j$ encapsulating a sensor pose, the a_{ji} form a unique row vector $\mathbf{a}_j^T = (a_{ji}) \in \mathbb{R}^n$ describing the coverage of each of the n basis functions when viewed from the respective sensor's perspective, and we get $m_j = \langle \mathbf{a}_j^T, \mathbf{x} \rangle$ in case of a good reconstruction $\mathbf{x}$.

Using all m measurements m_j, this leads to a linear system $A\mathbf{x} = \mathbf{m}$ with the *system matrix* $A = (a_{ji}) \in \mathbb{R}^{m \times n}$. A is generally not invertible, and to compute a reconstruction $\mathbf{x}$ we solve a least–squares problem instead (typically with iterative methods):

$$\min_{\mathbf{x}} \tfrac{1}{2} \|A\mathbf{x} - \mathbf{m}\|^2.$$

1.2   Quality of Acquisition Geometry

In devices using imaging gantries, the acquisition protocol is designed such that the solution of the linear system is as well–posed as possible. For instance in X–ray CT, the X–ray source/detector pair rotates around the region of interest Ω, and measurements are being collected on an arc–shaped trajectory of at least $180°$.

In case of freehand– or robot–guided data acquisition, however, the acquisition geometry is very sparsely and irregularly sampled, leading to very ill–posed problems. Fewer available measurements often result in under–determined systems, where a good sampling of the region of interest is mandatory. As communicating accurate instructions — particularly when orientation is involved — to a human operator is a complex and largely unsolved problem, we restrict ourselves to robot–controlled acquisitions in this work.

Ignoring application–specific approaches and focusing on a general setting, some of the approaches to characterize the ill–posedness of a system equation are:

1. *Column sums* [8]. Given m measurements, the linear system effectively contains m dot products between every row of the system matrix and the reconstruction $\mathbf{x}$. Vice versa, the i'th column of the system matrix contains all the m contributions of the basis function coefficient x_i to the measurements. A low value of the sum over all of column values is consequently an indicator for under–sampling, and enforcing maximal column sums yields better reconstructions.

2. *Singular spectrum of A* [5]. Over– and under–determined linear systems are solved via least squares methods, and the solution is for example obtained from the normal equation $A^T A \mathbf{x} = A^T \mathbf{m}$. In order for this system to have full rank, and thus a well–defined solution, the eigenvalues of $A^T A$ need to be of sufficiently large magnitude. This spectrum is exactly the set of singular values of A, and optimizing them accordingly during the acquisition improves the numerical condition of the system.

3. *Null–space of A* [9]. If $\mathbf{x}$ is a solution to $A\mathbf{x} = \mathbf{m}$, and if $\tilde{\mathbf{x}}$ is in the null–space of A, i.e. $\tilde{\mathbf{x}}$ is a solution to $A\tilde{\mathbf{x}} = \mathbf{0}$, then $\mathbf{x} + \tilde{\mathbf{x}}$ is also a solution of $A\mathbf{x} = \mathbf{m}$. To gain a unique solution for the inverse problem, the null–space has to be reduced to $\{\mathbf{0}\}$, which is typically not feasible. However, if all null–space vectors have a common sub–region of zero values, that region is uniquely determined by the acquisition geometry.

2   Methods

We generate a trajectory by iteratively selecting the best next sensor perspective out of a set of candidates. Such an approach requires two major components, generation of candidate perspectives, and quality estimation of each single candidate.

Using the notation from above, each sensor location is characterized by a unique row-vector $\mathbf{a}^T \in \mathbb{R}^n$. Describing the current state after k measurements by the system matrix $A \in \mathbb{R}^{k \times n}$ and the measurement vector $\mathbf{m} \in \mathbb{R}^k$, we are interested in finding an additional perspective, such that extending the linear system by the corresponding row $\mathbf{a}_{k+1}^T$ and measurement m_{k+1} yields an equation

$$A'\mathbf{x} = \mathbf{m}' \quad \text{or} \quad \begin{pmatrix} A \\ \mathbf{a}_{k+1}^T \end{pmatrix} \mathbf{x} = \begin{pmatrix} \mathbf{m} \\ m_{k+1} \end{pmatrix}$$

with a better quality estimate $\eta(\mathbf{a}^T)$.

2.1   SVD–Based Quality Estimation

In order to specify this aim accurately, we use a quality measure based on the singular value spectrum. Ideally, in comparison to A, the new system matrix A' will exhibit larger singular values. We enforce this objective by selecting new sensor poses that maximize the sum of singular values

$$\eta(\mathbf{a}^T) := \sum_{i=1}^{k+1} \sigma_i'$$

where $\sigma'_1 \geq \cdots \geq \sigma'_{k+1} \geq 0$ are the singular values of the extended matrix A', and $\mathbf{a}^T$ is constrained to be generated by a legal sensor perspective.

2.2 Incremental Computation of the SVD

Computing the singular values of a matrix is a numerically expensive operation, in particular considering the size of a typical system matrix. Due to the incremental nature of the system, it is reasonable to use a known decomposition of A when computing the factors of the extended matrix A' during the step $k \to k+1$. We focus on underdetermined systems ($k < n$), and use the economy-sized definition $A = UDV_1^T$, where $U \in \mathrm{SO}(k)$ is an orthogonal matrix, $D = \mathrm{diag}(\sigma_1, \ldots, \sigma_k) \in \mathbb{R}^{k \times k}$ a diagonal matrix holding the singular values $\sigma_1 \geq \cdots \geq \sigma_k \geq 0$, and $V_1 \in \mathbb{R}^{n \times k}$ a matrix with orthonormal columns.

Updating the SVD after adding a single row or column to a matrix is a problem already investigated in the fields of Data Mining, Latent Semantic Analysis, and also Computer Vision. There are several approaches, depending on whether exact values are required or whether approximations suffice. Also, some applications use the dominant singular values only, and omit the smaller ones entirely.

In our case, we are interested in the full spectrum of the squared system matrix $A^T A$, and we use the exact method presented by Gu and Eisenstat [3], including the optimizations proposed by Chetverikov and Axt [1]. Using $A = UDV_1^T$, we can preliminarily decompose A' into

$$A' = \begin{pmatrix} A \\ \mathbf{a}^T \end{pmatrix} = \underbrace{\begin{pmatrix} U & \mathbf{0} \\ \mathbf{0}^T & 1 \end{pmatrix}}_{=:\, M} \underbrace{\begin{pmatrix} D & \mathbf{0} \\ \mathbf{z}^T & \zeta \end{pmatrix}}_{=:\, L} \underbrace{\begin{pmatrix} V_1^T \\ \mathbf{v}^T \end{pmatrix}}_{=:\, N^T} = \begin{pmatrix} UDV_1^T \\ \mathbf{z}^T V_1^T + \zeta \mathbf{v}^T \end{pmatrix},$$

where $\mathbf{z} = V_1^T \mathbf{a} \in \mathbb{R}^k$ is the projection of $\mathbf{a}$ into the subspace defined by the rows of V_1. The other unknowns, $\zeta \in \mathbb{R}$ and $\mathbf{v} \in \mathbb{R}^n$, can be solved from the equation $\mathbf{w} := \mathbf{a} - V_1 V_1^T \mathbf{a} = \zeta \mathbf{v}$ which is obtained from the last row of the decomposition of A'. If the additional vector $\mathbf{a}^T$ is linearly independent of the rows of A, $\mathbf{v}$ is orthogonal to all columns of V_1, as required. Computing the SVD of the inner — relatively small — square matrix $L \in \mathbb{R}^{k \times k}$ using standard methods yields the decomposition $L = \widetilde{U}\widetilde{D}\widetilde{V}^T$ with both, $\widetilde{U}, \widetilde{V} \in \mathrm{SO}(k)$ orthogonal matrices. Using this and the preliminary decomposition $A' = MLN^T$, the economy-sized SVD of $A' = U'D'V_1'^T$ is given by $U' = M\widetilde{U}$, $D' = \widetilde{D}$, and $V' = N\widetilde{V}$.

We have extended this algorithm to support the addition of linearly dependent rows. This situation appears to be uncommon in other settings where real–world measurements are used, but may appear when searching for additional sensor poses in a structured way. In this case, $\|\mathbf{w}\| \approx 0$ and $\zeta \approx 0$, as expected, but $\mathbf{v}$ is usually no longer orthogonal to all columns of V_1 — leading to errors in the following step $k+1 \to k+2$. We detect this, and reinitialize $\mathbf{v}$ by creating an orthogonal vector by means of applying the Gram-Schmidt orthogonalization procedure to a random initial vector.

Please note that these equations are valid for underdetermined systems only, as we expect such a setting in our application of functional imaging. Equivalent

rules can be developed for all other linear systems in a very similar way, but we omit them here for brevity. The reader is kindly referred to the two original publications [3,1] for in–depth explanations and analyses.

2.3   Optimization and Trajectory Generation

Given a current system matrix A and a set of possible successor poses $C_{(1)}$, we can now select the best next pose out of that set using the techniques presented above. The remaining problem is that of generating that candidate set.

Given the singular value–based energy measure $\eta(\mathbf{a}^T)$, an optimal candidate row $\mathbf{a}^T$ will be as orthogonal as possible to the existing rows, thus optimizing the coverage of the basis functions b_i. Such a position is typically rather far away from the current pose, and a global search will be required to find it.

Identifying an orthogonal row, however, is a complex operation, but random selection has been shown to be a good replacement with high average success rate [7]. We consequently select candidate positions arbitrarily within a space of possible poses. The latter is used to impose constraints caused by geometric limitations, maximal measurement distances, etc.

The entire optimization procedure will thus start at an initial location given by row-vector $\mathbf{a}_{(0)}^T$ and the corresponding system matrix $A_{(0)} = (\mathbf{a}_{(0)}^T) \in \mathbb{R}^{1 \times n}$ with known SVD. With $C_{(1)}$ denoting the randomly selected candidate set, the second sensor perspective is chosen as

$$\mathbf{a}_{(1)}^T = \arg \max_{\mathbf{a}^T \in C_{(1)}} \eta(\mathbf{a}^T),$$

yielding an extended system matrix $A_{(1)} \in \mathbb{R}^{2 \times n}$ with maximal sum over the singular value spectrum. That process is repeated until the required number of poses has been reached.

Such a random path will obviously contain large hops, and to minimize the acquisition path length and time to scan it is essential to post–process it. We use a two–stage approach to sort the poses into a useful sequence, first partitioning all positions into local clusters, and then reordering each of them individually using an approximative Traveling–Salesman–solver based on the Minimum Spanning Tree heuristic. If steps of large size remain, we insert 'safe positions', and recombine the partitions, yielding a smooth acquisition path that can for example be traced by a robot in reasonable time.

3   Experiments and Results

3.1   Experiments

In order to test the optimization procedure, we created a Matlab script computing trajectories. Path planning has been performed on a laptop computer with an Intel Core i7 CPU with 4 processing cores at 2.3 GHz and 8 GB of memory.

The experiments are based on an intra–operative functional imaging setup similar to freehand SPECT [8], using a tracked gamma detector to generate

localized reconstruction of a radioactivity distribution. Assuming a region of interest surrounded by a bounding box of $(10 \text{ cm})^2 \times 5$ cm, we used two alternative parameterizations of two degrees of freedom, generating locations on a hemisphere around the base plane's center, then retracting the sensor along that direction to reach the surface of the bounding box. The first parameterization used spherical coordinates, the other directly created unit directions in Cartesian space. Of the six sides of the cuboid, we only measure on the top plane, as well as on two orthogonal side planes, thus imitating the spatial constraints of an intra–operative situation where patient body and operating bed place constraints on accessibility.

Using a rather coarse discretization of that region of $10 \times 10 \times 5$ voxels, we were able to generate a path of 300 positions in about 140 seconds. The coarse discretization is reasonable as the small system matrix shows similar behavior as observed when working with its full–size equivalent. Furthermore, when using a robot to traverse the trajectory, additional data can be recorded during the movements to record considerably more than just 300 measurements.

(a) Energies over path length   (b) Partitions and trajectory
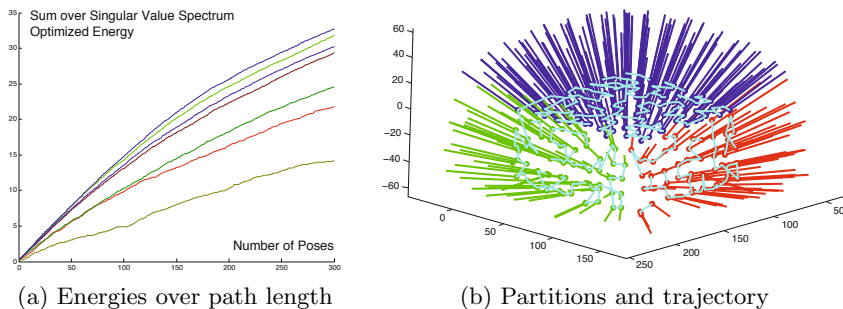
Fig. 1. Evaluation. (1a) Energy curves denote, from bottom up, a human measurement, random sampling with 1 candidate and spherical coordinates (SC), with 1 candidate and Cartesian coordinates (CC), 5 candidates SC, 5 candidates CC, 12 candidates SC, 12 candidates CC. (1b) Sampling of the (invisible) region of interest; acquisition locations on three bounding box planes (red/green/blue = left/front/top), and robot trajectory (cyan).

3.2   Results

Singular-Values-Spectrum over Path Length In a first experiment, we compared the energy $\eta(\mathbf{a}^T)$ at different evolution stages for different sampling strategies. The result is shown in figure (1a). We used candidate sets of size 1, 5, and 12 and the two mentioned parameterizations. In general, more samples yielded better results, and sampling in Cartesian space turned out to be slightly superior — this parameterization does not show any clustering, as can be observed around the poles when using spherical coordinates.

We also evaluated the evolution of the energy $\eta(\mathbf{a}^T)$ for a trajectory recorded while a human performed a standard protocol. In terms of our energy, this path

is the worst, potentially due to the fact that the probe is primarily translated, while the orientation vector remains almost static for a considerable amount of time. Also, the number of measurements per side is not proportional to the surface area.

Simulated Measurements. Next, we generated ground truth volumes and simulated measurements, and compared the reconstruction results for several trajectories, generated by humans and the optimization procedure. The reconstructions were considerably underdetermined (300 measurements for 500,000 unknowns), solved by MLEM [4] (20 iterations), and show SPECT–typical displacement errors. Nevertheless, the reconstructions based on trajectories created by the proposed method show better separation between the hot spots, and considerably less 'activity bleeding'. Examples are shown in figures (2a) and (2b).

Real Robot-Guided Measurements. Finally, we fed our trajectories to a robot arm guiding the gamma detector. An example image showing the partitioned sets of probe positions, and the robot trajectory (omitting the intermediate safe positions) is shown in figure (1b). Results are given in figures (2c), (2d).

(a) Human   (b) SVD-based   (c) Robot 1   (d) Robot 2
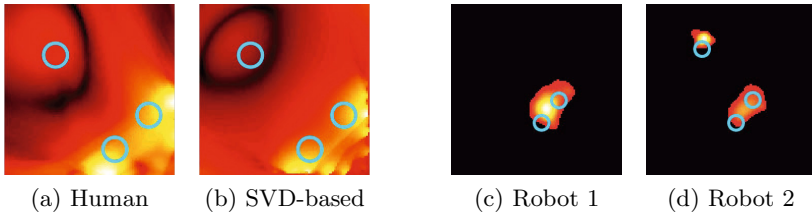
Fig. 2. Reconstruction results, looking top–down, with blue circles denoting ground truth locations of activity seeds. (2a)–(2b) show the logarithms of the per–column standard deviations of the intensities in a *simulated* setting. The path generated by our method yields a better separation and improved circumscription of the hot regions. (2c)–(2d) show the results of two *real* acquisitions performed by a robot following a trajectory created by our method.

4   Discussion and Conclusion

We have presented a method to generate an optimized trajectory for tomographic reconstruction in intra–operative settings. This optimality is defined based on the singular value spectrum, and the corresponding measure is computed using fast incremental updates of the SVD. Since this only depends on the system matrix, this approach is applicable for any imaging modality. Even though our experiments used a regular bounding box, more complicated geometries given as a polygon mesh, for instance obtained by laser–scanning the patient, could also be used to generate trajectories.

Several improvements can be made. Runtime can be improved by an efficient implementation as well as by exploiting the 'broken–arrowhead' structure of the inner matrix while computing its SVD, thus enabling the computation of longer trajectories. Furthermore, it would be interesting to investigate other energy measures based on the kernel or column sums.

The most interesting point will be to convert this approach to a real–time path planning application considering the kinematics of a robot. While the energy measure is sufficiently fast, a naïve local search for candidates is prone to get stuck at local maxima. A strategic planner will need to consider now what other positions to visit later, while still guaranteeing both full coverage and smooth motion.

References

1. Chetverikov, D., Axt, A.: Approximation-free Running SVD and its Application to Motion Detection. Pattern Recognition Letters 31(9), 891–897 (2010)
2. Ganguly, A., Fieselmann, A., Marks, M., Rosenberg, J., Boese, J., Deuerling-Zheng, Y., Straka, M., Zaharchuck, G., Bammer, R., Fahrig, R.: Cerebral CT Perfusion Using an Interventional C-Arm Imaging System: Cerebral Blood Flow Measurements. American Journal of Neuroradiology 32, 1525–1531 (2011)
3. Gu, M., Eisenstat, S.C.: A Stable and Fast Algorithm for Updating the Singular Value Decomposition. Research Report RR-966, Yale University, New Haven (1994)
4. Herman, G.T.: Fundamentals of Computerized Tomography, 2nd edn. Springer, London (2010)
5. Lasser, T., Ntziachristos, V.: Optimization of 360° Projection Fluorescence Molecular Tomography. Medical Image Analysis 11(4), 389–399 (2007)
6. Schneider, C., Guerrero, J., Nguan, C., Rohling, R., Salcudean, S.: Intra-operative "Pick-Up" Ultrasound for Robot Assisted Surgery with Vessel Extraction and Registration: A Feasibility Study. In: Taylor, R.H., Yang, G.-Z. (eds.) IPCAI 2011. LNCS, vol. 6689, pp. 122–132. Springer, Heidelberg (2011)
7. Strohmer, T., Vershynin, R.: A Randomized Solver for Linear Systems with Exponential Convergence. In: Díaz, J., Jansen, K., Rolim, J.D.P., Zwick, U. (eds.) APPROX and RANDOM 2006. LNCS, vol. 4110, pp. 499–507. Springer, Heidelberg (2006)
8. Wendler, T., Herrmann, K., Schnelzer, A., Lasser, T., Traub, J., Kutter, O., Ehlerding, A., Scheidhauer, K., Schuster, T., Kiechle, M., Schwaiger, M., Navab, N., Ziegler, S., Buck, A.: First demonstration of 3-D lymphatic mapping in breast cancer using freehand SPECT. European Journal of Nuclear Medicine and Molecular Imaging 37(8), 1452–1461 (2010)
9. Zeng, G.L., Gullberg, G.T.: Null-Space Function Estimation for the Three-Dimensional Interior Problem. In: Proc. of Fully 3D, Potsdam, pp. 241–245 (2011)

Incorporating Parameter Uncertainty in Bayesian Segmentation Models: Application to Hippocampal Subfield Volumetry

Juan Eugenio Iglesias[1], Mert Rory Sabuncu[1], Koen Van Leemput[1,2,3], and The Alzheimer's Disease Neuroimaging Initiative

[1] Martinos Center for Biomedical Imaging, MGH, Harvard Medical School, USA
[2] Department of Informatics and Mathematical Modeling, DTU, Denmark
[3] Departments of Information and Computer Science and of Biomedical Engineering and Computational Science, Aalto University, Finland

Abstract. Many successful segmentation algorithms are based on Bayesian models in which prior anatomical knowledge is combined with the available image information. However, these methods typically have many free parameters that are estimated to obtain point estimates only, whereas a faithful Bayesian analysis would also consider all possible alternate values these parameters may take. In this paper, we propose to incorporate the uncertainty of the free parameters in Bayesian segmentation models more accurately by using Monte Carlo sampling. We demonstrate our technique by sampling atlas warps in a recent method for hippocampal subfield segmentation, and show a significant improvement in an Alzheimer's disease classification task. As an additional benefit, the method also yields informative "error bars" on the segmentation results for each of the individual sub-structures.

1 Introduction

Many segmentation algorithms in medical image analysis are based on Bayesian modeling, in which generative image models are constructed and subsequently "inverted" to obtain automated segmentations. Such methods have a *prior* that makes predictions about where anatomical structures typically occur throughout the image, such as Markov random field models or probabilistic atlases [1, 2]. They also include a *likelihood* term that models the relationship between segmentation labels and image intensities, often incorporating explicit models of imaging artifacts [3]. Once the prior and likelihood have been specified, the posterior distribution over all possible segmentations can be inferred using Bayes' rule. It is then possible to search for the segmentation that maximizes this posterior, or to directly estimate from it the volumes of specific structures.

Although these methods are clearly "Bayesian", an issue that is usually overlooked is that they only apply Bayesian analysis in an *approximate* sense. In particular, these models typically have many free parameters for which suitable values are unknown a priori. In a true Bayesian approach, such parameters need

to be integrated over when inferring the segmentation posterior. But, in practice, their *optimal* values are first estimated and only the resulting point estimates are used to compute the segmentation posterior instead. In recent years generative models have started to include deformable registration methods that warp probabilistic atlases into the domain of the image being analyzed, often adding thousands of free parameters to the model [4–7]. Since many plausible atlas warps beside the truly optimal one may exist, computing segmentations based on a single warp may lead to biased results. Furthermore, the numerical optimizers computing such high-dimensional atlas warps may not necessarily find the global optimum, further contributing to segmentation errors.

In this paper, we investigate the effect of using a more accurate approximation of the segmentation posterior in Bayesian segmentation models than the point estimates of the free model parameters. In particular, we will approximate the integral over atlas deformations in a recently proposed method for hippocampal subfield segmentation [7] using Markov chain Monte Carlo (MCMC) sampling, and compare the results to those obtained using the most probable warp only. We show that MCMC sampling yields hippocampal subfield volume estimates that better discriminate controls from subjects with Alzheimer's disease, while providing informative "error bars" on those estimates as well.

To the best of our knowledge, the issue of integrating over free parameters in Bayesian segmentation models has not been addressed before in the literature. The closest work related to the techniques used in this paper infers the posterior distribution of deformation fields in the context of computing location-specific smoothing kernels [8], quantifying registration uncertainties [9], or constructing Bayesian deformable models [10].

2 Methods

2.1 Baseline Segmentation Method

We start from the Bayesian method for hippocampal subfield segmentation [7] that is publicly available as part of the FreeSurfer software package[1]. In this method, a segmentation prior is defined in the form of a tetrahedral mesh-based probabilistic atlas in which each mesh vertex has an associated vector of probabilities for the different hippocampal subfields and surrounding tissues (fimbria, presubiculum, subiculum, CA1, CA2/3, CA4/DG, hippocampal fissure, white matter, gray matter, and CSF). The resolution and topology of the mesh are locally adaptive to the level of shape complexity of each anatomical region, e.g., it is coarse in uniform regions and fine around convoluted boundaries. The mesh can be deformed according to a probabilistic model on the location of the mesh nodes $p(\mathbf{x}) \propto \exp(-\phi(\mathbf{x}))$, where $\mathbf{x}$ is a vector containing the coordinates of the mesh nodes, and $\phi(\mathbf{x})$ is an energy function that penalizes mesh positions in which the tetrahedra are deformed [11]. This function goes to infinity if the Jacobian determinant of any tetrahedron's deformation approaches zero, and

---

[1] http://surfer.nmr.mgh.harvard.edu/

therefore ensures that the mesh topology is preserved. For a given $\mathbf{x}$, the prior probability $p_i(k|\mathbf{x})$ of tissue k occurring in voxel i is obtained by interpolating the probability vectors in the vertices of the deformed mesh. Assuming conditional independence of the labels between voxels given $\mathbf{x}$, the prior probability of a segmentation is then given by $p(\mathbf{l}|\mathbf{x}) = \prod_i p_i(l_i|\mathbf{x})$, where $\mathbf{l} = (l_1, \ldots, l_I)^{\mathrm{T}}$, $l_i \in \{1, \ldots, K\}$ is a segmentation of an image with I voxels into K tissue types.

For the likelihood, we model the intensity of voxels in tissue k as a Gaussian distribution with parameters μ_k, σ_k^2: $p(\mathbf{y}|\mathbf{l}, \boldsymbol{\theta}) = \prod_i \mathcal{N}(y_i; \mu_{l_i}, \sigma_{l_i}^2)$, where the vector $\mathbf{y} = (y_1, \ldots, y_I)^T$ contains the image intensities, and $\boldsymbol{\theta} = (\mu_1, \sigma_1^2, \ldots, \mu_K, \sigma_K^2)^T$ represents the Gaussian distribution parameters. A non-informative prior for $\boldsymbol{\theta}$ (i.e., $p(\boldsymbol{\theta}) \propto 1$) completes the model.

Given an image to segment, the posterior over possible segmentations is given by $p(\mathbf{l}|\mathbf{y}) = \int_{\boldsymbol{\theta}} \int_{\mathbf{x}} p(\mathbf{l}|\mathbf{y}, \mathbf{x}, \boldsymbol{\theta}) p(\mathbf{x}, \boldsymbol{\theta}|\mathbf{y}) \mathrm{d}\mathbf{x}\mathrm{d}\boldsymbol{\theta}$, which takes into account the contribution of all possible values for the model parameters $\{\mathbf{x}, \boldsymbol{\theta}\}$, each weighted by their posterior probability $p(\mathbf{x}, \boldsymbol{\theta}|\mathbf{y})$. In [7], this integral is approximated by estimating the parameters with maximal weight $\{\hat{\mathbf{x}}, \hat{\boldsymbol{\theta}}\} = \arg\max_{\{\mathbf{x}, \boldsymbol{\theta}\}} p(\mathbf{x}, \boldsymbol{\theta}|\mathbf{y})$, and using the contribution of those parameters only, yielding

$$p(\mathbf{l}|\mathbf{y}) \simeq p(\mathbf{l}|\mathbf{y}, \hat{\mathbf{x}}, \hat{\boldsymbol{\theta}}) = \prod_i p_i(l_i|y_i, \hat{\mathbf{x}}, \hat{\boldsymbol{\theta}}) \tag{1}$$

$$\text{with} \quad p_i(k|y_i, \hat{\mathbf{x}}, \hat{\boldsymbol{\theta}}) \propto \mathcal{N}(y_i; \hat{\mu}_k, \hat{\sigma}_k^2) p_i(k|\hat{\mathbf{x}}). \tag{2}$$

The segmentation maximizing this approximate posterior is obtained by simply assigning each voxel to the tissue class that maximizes Eq. (2). Furthermore, the volume of class k also has an (approximate) posterior distribution, with mean

$$v_k = \sum_i p_i(k|y_i, \hat{\mathbf{x}}, \hat{\boldsymbol{\theta}}) \tag{3}$$

and variance

$$\gamma_k^2 = \sum_i p_i(k|y_i, \hat{\mathbf{x}}, \hat{\boldsymbol{\theta}})[1 - p_i(k|y_i, \hat{\mathbf{x}}, \hat{\boldsymbol{\theta}})]. \tag{4}$$

2.2   Incorporating Parameter Uncertainty

The approximation of Eq. (1) will be a good one if the posterior of the model parameters, $p(\mathbf{x}, \boldsymbol{\theta}|\mathbf{y})$, is very peaked around $\{\hat{\mathbf{x}}, \hat{\boldsymbol{\theta}}\}$. Although this is a reasonable assumption for the Gaussian distribution parameters $\boldsymbol{\theta}$ – one cannot alter them much without considerably decreasing the likelihood of the model – assuming a sharp peak for the mesh position $\mathbf{x}$ is not necessarily accurate, since moving vertices in areas with low image contrast does not drastically change $p(\mathbf{x}, \boldsymbol{\theta}|\mathbf{y})$.

We therefore propose to use a computationally more demanding but more accurate way of approximating $p(\mathbf{l}|\mathbf{y})$. Specifically, we propose to draw a number of samples $\mathbf{x}(n)$, $n = 1, \ldots, N$ from the posterior distribution $p(\mathbf{x}|\mathbf{y}, \hat{\boldsymbol{\theta}})$ using Monte Carlo sampling, and approximate the segmentation posterior by

$$p(\mathbf{l}|\mathbf{y}) \; \simeq \; \int_{\mathbf{x}} p(\mathbf{l}|\mathbf{y}, \mathbf{x}, \hat{\boldsymbol{\theta}}) p(\mathbf{x}|\mathbf{y}, \hat{\boldsymbol{\theta}}) \mathrm{d}\mathbf{x} \; \simeq \; \frac{1}{N} \sum_{n=1}^{N} p(\mathbf{l}|\mathbf{y}, \mathbf{x}(n), \hat{\boldsymbol{\theta}}), \tag{5}$$

where in the first step we have used the mode approximation in the direction of $\boldsymbol{\theta}$, as before, but in the second step the remaining integral is approximated by summing the contributions of many possible atlas warps (with more probable warps occurring more frequently), rather than by the contribution of a single point estimate $\hat{\mathbf{x}}$ only. Given enough samples, this approximation can be made arbitrarily close to the true integral.

Once N samples $\mathbf{x}(n)$ are available, it follows from Eqs. (3–5) that the approximate posterior for the volume of tissue class k has mean and variance

$$v_k = \frac{1}{N} \sum_{n=1}^{N} v_k(n) \tag{6}$$

$$\gamma_k^2 = \frac{1}{N} \left[\sum_{n=1}^{N} [v_k(n) - v_k]^2 + \gamma_k^2(n) \right], \tag{7}$$

respectively, where $v_k(n) = \sum_i p_i(k|y_i, \mathbf{x}(n), \hat{\boldsymbol{\theta}})$ and $\gamma_k^2(n) = \sum_i p_i(k|y_i, \mathbf{x}(n), \hat{\boldsymbol{\theta}})$ $[1 - p_i(k|y_i, \mathbf{x}(n), \hat{\boldsymbol{\theta}})]$.

2.3 MCMC Sampling

In order to obtain the required samples $\mathbf{x}(n)$, we use a MCMC sampling technique known as the Hamiltonian Monte Carlo (HMC) method [12], which is more efficient than traditional Metropolis schemes because it uses gradient information to reduce random walk behavior. Specifically, it facilitates large steps in $\mathbf{x}$ with relatively few evaluations of the target distribution $p(\mathbf{x}|\mathbf{y}, \hat{\boldsymbol{\theta}})$ and its gradient, by iteratively assigning a random momentum to each component of $\mathbf{x}$, and then simulating the Hamiltonian dynamics of a system in which $-\log p(\mathbf{x}|\mathbf{y}, \hat{\boldsymbol{\theta}})$ acts as an internal "force". In our implementation, we discretize the Hamiltonian trajectories using the so-called leapfrog method [12], and simulate the Hamiltonian dynamics for a number of time steps sampled uniformly from $[1, 50]$ to obtain a proposal for the Metropolis algorithm. Discretization step sizes that are adequate for some tetrahedra might be too large or small for others, leading to either slow convergence or too many rejected moves. We therefore use the following heuristic stepsize for each vertex: $\eta/\max[\partial^2(-\log p(\mathbf{x}))/\partial \mathbf{x}_j^2|_{\hat{\mathbf{x}}}]$, where η is a global adjustment factor and $\partial^2/\mathbf{x}_j^2$ denotes the second derivatives with respect to the three spatial coordinates of vertex j. Two samples of $p(\mathbf{x}|\mathbf{y}, \hat{\boldsymbol{\theta}})$ obtained using the proposed scheme are displayed in Fig. 1.

3 Experiments and Results

To investigate the effect of approximating the true posterior over the segmentations using parameter sampling instead of point estimates, we compared the performance of the estimated subfield volumes for both methods (Eq. (3) vs. Eq. (6)) in an Alzheimer's disease classification task[2]. In particular, we collected the

---

[2] Although this specific classification task is best performed using information from the whole brain [13], the goal of this paper is to show the effect of MCMC sampling.

volume estimates for all seven subfields (averaged over the left and right hemispheres) into a feature vector **v** for each subject, and trained and tested a simple multivariate classifier to discern between elderly controls (EC) and Alzheimer's disease patients (AD) in the corresponding feature space. We also compared the variance ("error bars") on the subfield volume estimates for both methods (Eq. (4) vs. Eq. (7)), and investigated the effect of incorporating this information in the training of the classifier as well.

3.1 Data and Experimental Set-Up

The 400 baseline T_1 scans from controls and AD subjects available in ADNI [3] where used in this study. The MRI pulse sequence is described elsewhere[3]. The volumes were preprocessed and parsed into 36 brain structures using FreeSurfer. We discarded 17 subjects for which FreeSurfer crashed. The demographics for the remaining 383 were: 56.2% controls (age 76.1 ± 5.6), 43.8% Alzheimer's (age 75.5 ± 7.6); 53.6% males (age 76.1 ± 5.6), 46.4% females (age 75.9 ± 6.8).

After the segmentation of subcortical structures, the FreeSurfer hippocampal subfield segmentation routine (Section 2.1) was executed. The output $\{\hat{\mathbf{x}}, \hat{\boldsymbol{\theta}}\}$ was used to initialize the HMC sampler, which was then used to generate $N = 50$ samples per subject. The parameter η was tuned so that the average Metropolis rejection rate was approximately 25%. To decrease the correlation between successive samples, we recorded **x** at the end of every $200th$ Hamiltonian trajectory (chosen by visual inspection of the autocorrelation of subsequent runs). We allowed 300 initial "burn-in" runs before collecting samples. The running time of the sampling was roughly three hours.

3.2 Classification and ROC Analysis

We used a Quadratic Discriminant Analysis (QDA) classifier, which assumes that the feature vectors **v** in each group are normally distributed according to $\mathcal{N}(\mathbf{v}|\boldsymbol{\mu}_{EC}, \boldsymbol{\Sigma}_{EC})$ and $\mathcal{N}(\mathbf{v}|\boldsymbol{\mu}_{AD}, \boldsymbol{\Sigma}_{AD})$, respectively. The means and covariances were estimated from the available training samples. In testing, a subject was classified as EC or AD by thresholding the likelihood ratio $\mathcal{N}(\mathbf{v}|\boldsymbol{\mu}_{EC}, \boldsymbol{\Sigma}_{EC})/\mathcal{N}(\mathbf{v}|\boldsymbol{\mu}_{AD}, \boldsymbol{\Sigma}_{AD}) \lessgtr \lambda$. The corresponding ROC curve (i.e., true positive rate vs. false positive rate) was obtained by sweeping the threshold λ, and the area under the curve (A_z) was then used as a measure of performance. The ROCs were computed using cross-validation with two randomly selected folds.

We also analyzed the accuracy when the volume of the whole hippocampus is thresholded to separate EC from AD. We compared two estimates of the volume: (1) the sum of the volumes of the subfields; and (2) the estimate from the FreeSurfer pipeline. Finally, to assess the effect of sampling on training and testing separately, we conducted an experiment in which the classifier was trained on point estimate volumes and evaluated on MCMC volumes, and vice versa.

---

[3] Online at http://www.adni-info.org/

3.3   Results

Fig. 2 shows the ROC curves and the areas under them (A_z) for the different methods. Also shown are the p-values of paired DeLong statistical tests [14] that evaluate if the differences in A_z are significant. At $p = 0.05$, sampling significantly outperformed point estimates in all cases (subfields and whole hippocampus). At the operating point closest to $(0, 1)$, sampling provides a $\sim 2\%$ increase in classification accuracy. Using all the subfields performed significantly better than the whole hippocampal volume alone. All methods based on the subfield analysis outperformed the standard FreeSurfer hippocampal segmentation.

When the QDA was trained on the point estimate subfield volumes and tested on those obtained with sampling, we obtained $A_z = 0.875$, and when the roles were switched, $A_z = 0.876$. These values are better than when point estimate volumes were used for both training and testing, but worse than when sampling was used throughout, indicating that MCMC sampling is beneficial for both obtaining better discriminative directions and classifying individual subjects.

We also compared the variances of the hippocampal subfield volume posteriors (Table 1). The point estimates (Eq. (4)) clearly underestimate them, especially for the larger subfields; e.g., the standard error for CA2-3 is 0.4% of its volume, unrealistic given the poor image contrast (Fig. 1). In contrast, sampling (Eq. (7)) produces values between 5% and 10%, better reflecting the uncertainty in the estimated volumes.

In an attempt to take the MCMC volumetry uncertainty estimates into account in the classifier, we also trained a QDA by simply using all contributing volumes $v_k(n), n = 1, \ldots, N = 50$ in Eq. (6) for each subject – effectively using 50 times more training samples than there are training subjects. The ROC and the corresponding A_z are displayed in Fig. 2 (labeled as "error bars"), showing a modest further improvement compared to when the classifier is trained using the mean values only. Although the improvement was not statistically significant ($p \approx 0.1$), the ROC seems to be consistently better in the region that is closest to (0,1), where the operating point of the classifier would be typically defined.

4   Discussion

In this paper we proposed to approximate the segmentation posterior in probabilistic segmentation models more faithfully by using Monte Carlo samples of their free parameters. We demonstrated our technique by sampling atlas warps in a Bayesian method for hippocampal subfield segmentation, and showed a significant improvement in an Alzheimer's disease classification task. The method is general and can also be applied to other Bayesian segmentation models. It yields realistic confidence intervals on the segmentation results of individual structures, which we believe will convey important information when these techniques are ultimately applied in clinical settings. Furthermore, such confidence information may also help select the most suitable scanning protocol for imaging studies investigating the morphometry of specific anatomical structures.
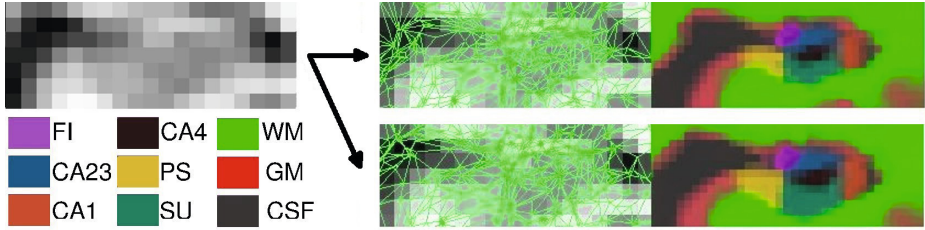
Fig. 1. A coronal slice of an MR scan, zoomed in around the right hippocampus, and two different samples from $p(\mathbf{x}|\mathbf{y}, \hat{\boldsymbol{\theta}})$. Left: deformed mesh; right: corresponding priors $p(\mathbf{l}|\mathbf{x})$ (at the locations in which more than one class prior is greater than zero, the color is a linear combination of the class colors, weighted by their corresponding probabilities). The abbreviations in the color code are: FI: fimbria, PS: presubiculum, SU: subiculum, WM: white matter, GM: gray matter.

Table 1. Mean volumes and relative standard deviations (γ_k/v_k) for the different subfields, estimated using point estimates and MCMC samples of atlas deformations. HF stands for "hippocampal fissure"; the other abbreviations are as in Fig. 2.

Subfield	HF	FI	CA4	CA1	PS	SU	CA23
Volume (mm^3)	38	56	248	265	324	326	517
γ_k/v_k, point est. (%)	5.5	1.0	1.2	0.5	0.3	0.7	0.4
γ_k/v_k, sampling (%)	9.9	4.8	7.3	7.4	6.3	8.0	5.8
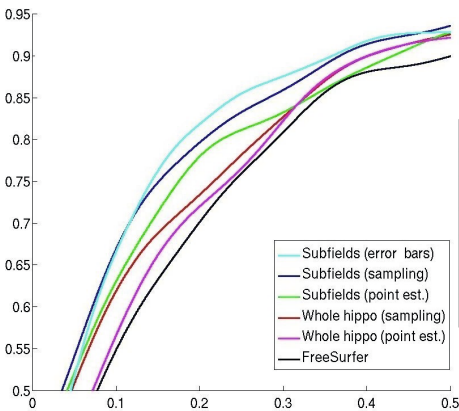
Method	A_z	$p_{SF_{sp}}$	$p_{SF_{pe}}$	$p_{WH_{sp}}$	$p_{WH_{pe}}$	p_{FS}
SF_{eb}	.884	.101	.022	.011	.005	.002
SF_{sp}	.882	-	.031	.019	.010	.006
SF_{pe}	.869	-	-	.046	.019	.009
WH_{sp}	.857	-	-	-	.032	.012
WH_{pe}	.844	-	-	-	-	.025
FS	.834	-	-	-	-	-

Fig. 2. Left: ROC curves for the different methods. "FreeSurfer" refers to the whole hippocampus segmentation produced using the standard FreeSurfer pipeline. Note that only the region $[0, 0.5] \times [0.5, 0.95]$ is shown. Right: Area under the curve (A_z) for each method as well as p-values corresponding to DeLong tests comparing A_z for different methods. "SF" stands for subfields, "WH" for whole hippocampus, "pe" for point estimate, "sp" for sampling, "eb" for sampling with error bars (i.e. using all volumes $v_k(n)$ in Eq. (6)), and "FS" for FreeSurfer.

References

1. Zhang, Y., Brady, M., Smith, S.: Segmentation of brain MR images through a hidden Markov random field model and the Expectation-Maximization algorithm. IEEE Transactions on Medical Imaging 20(1), 45–57 (2001)
2. Fischl, B., Salat, D., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., Dale, A.: Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. Neuron 33, 341–355 (2002)
3. Wells, W., Grimson, W., Kikinis, R., Jolesz, F.: Adaptive segmentation of MRI data. IEEE Transactions on Medical Imaging 15(4), 429–442 (1996)
4. Fischl, B., Salat, D., van der Kouwe, A., Makris, N., Segonne, F., Quinn, B., Dale, A.: Sequence-independent segmentation of magnetic resonance images. NeuroImage 23, S69–S84 (2004)
5. Ashburner, J., Friston, K.: Unified segmentation. NeuroImage 26, 839–851 (2005)
6. Pohl, K., Fisher, J., Grimson, W., Kikinis, R., Wells, W.: A Bayesian model for joint segmentation and registration. NeuroImage 31(1), 228–239 (2006)
7. Van Leemput, K., Bakkour, A., Benner, T., Wiggins, G., Wald, L., Augustinack, J., Dickerson, B., Golland, P., Fischl, B.: Automated segmentation of hippocampal subfields from ultra-high resolution in vivo MRI. Hippocampus 19, 549–557 (2009)
8. Simpson, I.J.A., Woolrich, M., Groves, A.R., Schnabel, J.A.: Longitudinal Brain MRI Analysis with Uncertain Registration. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part II. LNCS, vol. 6892, pp. 647–654. Springer, Heidelberg (2011)
9. Risholm, P., Pieper, S., Samset, E., Wells III, W.M.: Summarizing and Visualizing Uncertainty in Non-rigid Registration. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part II. LNCS, vol. 6362, pp. 554–561. Springer, Heidelberg (2010)
10. Allassonniére, S., Amit, Y., Trouvé, A.: Toward a coherent statistical framework for dense deformable template estimation. Journal of the Royal Statistical Society, Series B 69, 3–29 (2007)
11. Ashburner, J., Andersson, J., Friston, K.: Image registration using a symmetric prior – in three dimensions. Human Brain Mapping 9(4), 212–225 (2000)
12. Duane, S., Kennedy, A., Pendleton, B., Roweth, D.: Hybrid Monte Carlo. Physics Letters B 195(2), 216–222 (1987)
13. Cuingnet, R., Gerardin, E., Tessieras, J., Auzias, G., Lehéricy, S., Habert, M., 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 56(2), 766–781 (2011)
14. DeLong, E., DeLong, D., Clarke-Pearson, D.: Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics, 837–845 (1988)

A Dynamical Appearance Model Based on Multiscale Sparse Representation: Segmentation of the Left Ventricle from 4D Echocardiography[⋆]

Xiaojie Huang[1], Donald P. Dione[4], Colin B. Compas[2],
Xenophon Papademetris[2,3], Ben A. Lin[4], Albert J. Sinusas[3,4],
and James S. Duncan[1,2,3]

Departments of
[1] Electrical Engineering
[2] Biomedical Engineering
[3] Diagnostic Radiology
[4] Internal Medicine,
Yale University, New Haven, CT, USA
`xiaojie.huang@yale.edu`

Abstract. The spatio-temporal coherence in data plays an important role in echocardiographic segmentation. While learning offline dynamical priors from databases has received considerable attention, these priors may not be suitable for post-infarct patients and children with congenital heart disease. This paper presents a dynamical appearance model (DAM) driven by individual inherent data coherence. It employs multiscale sparse representation of local appearance, learns online multiscale appearance dictionaries as the image sequence is segmented sequentially, and integrates a spectrum of complementary multiscale appearance information including intensity, multiscale local appearance, and dynamical shape predictions. It overcomes the limitations of database-driven statistical models and applies to a broader range of subjects. Results on 26 4D canine echocardiographic images acquired from both healthy and post-infarct subjects show that our method significantly improves segmentation accuracy and robustness compared to a conventional intensity model and our previous single-scale sparse representation method.

1 Introduction

Segmentation of the left ventricle from 4D echocardiography plays an essential role in quantitative cardiac functional analysis. Due to gross image inhomogeneities, artifacts, and poor contrast between regions of interest, robust and accurate automatic segmentation of the left ventricle, especially the epicardial border, is very challenging in echocardiography. The inherent spatio-temporal coherence of echocardiographic data provides important constraints that can be exploited to guide cardiac border estimation and has motivated a spatio-temporal view point of echocardiographic segmentation.

Following the seminal work of Cootes et al. [1] on statistical shape/appearance modeling, a number of spatio-temporal statistical models (e.g., [2–6]) have been proposed for learning dynamical priors offline from databases. While these models have advantages in different aspects, the problem of forming a database that can handle a wide range of normal and abnormal heart images is still open to our knowledge. The assumption that different subjects have similar shape or motion pattern or their clinical images have similar appearance may not hold for routine clinical images, especially for disease cases, due to natural subject-to-subject tissue property variations and operator-to-operator variation in acquisition [7]. For example, for post-infarct patients, the positions, sizes and shapes of infarcts and thereby the overall heart motion can be highly variable across the population. It is very hard to build a reliable database accounting for all these variations, while such individual uniqueness is essentially desired information in some important applications like motion-based functional analysis. In addition, the tremendous cost of building reliable databases compromises the attractiveness of the database-driven methods.

Exploiting individual data coherence through online learning overcomes these limitations. It is particularly attractive when a database is inapplicable, unavailable, or defective. To this end, a model is indispensable for reliably uncovering the inherent spatio-temporal structure of individual 4D data. Sparse representation is a powerful mathematical framework for studying high-dimensional data. We proposed a 2D single-scale sparse-representation-based segmentation method in [8]. It shows the feasibility of analyzing 2D+t echocardiographic images via sparse representation and online dictionary learning. However, it is difficult to directly apply this method to 4D data. An important limitation is that it utilizes only a single scale of appearance information and requires careful tuning of scale parameters. This compromises segmentation accuracy and robustness. This paper generalizes our previous work [8] and introduces a new 3D dynamical appearance model (DAM) that leverages a full spectrum of complementary multiscale appearance information including intensity, multiscale local appearance, and shape. It employs multiscale sparse representation of high-dimensional local appearance, encodes appearance patterns with multiscale appearance dictionaries, and dynamically updates the dictionaries as the frames are segmented sequentially. The online multiscale dictionary learning process is supervised in a boosting framework to seek optimal weighting of multiscale information and generate dictionaries that are both generative and discriminative. Sparse coding w.r.t. the predictive dictionaries produces a local appearance discriminant. We also include intensity and a dynamical shape prediction to complete the appearance spectrum that we incorporate into a MAP framework.

2   Methods

2.1   Multiscale Sparse Representation

Let Ω denote the 3D image domain. We describe the multiscale local appearance at a pixel $\mathbf{u} \in \Omega$ in frame I_t with a series of appearance vectors $\mathbf{y}_t^k(\mathbf{u}) \in \mathbb{R}^n$

at different appearance scales $k = 1, ..., J$. $\mathbf{y}_t^k(\mathbf{u})$ is constructed by concatenating orderly the pixels within a block centered at $\mathbf{u}$. Complementary multiscale appearance information is extracted using a fixed block size at different levels of Gaussian pyramid. Modeled with sparse representation, an appearance vector $\mathbf{y} \in \mathbb{R}^n$ can be represented as a sparse linear combination of the atoms from an appearance dictionary $\mathbf{D} \in \mathbb{R}^{n \times K}$ which encodes the typical patterns of a corresponding appearance class. That is, $\mathbf{y} \approx \mathbf{Dx}$. Given $\mathbf{y}$, $\mathbf{D}$, and a sparsity factor T_0, the sparse representation $\mathbf{x}$ can be solved by sparse coding:

$$\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{Dx}\|_2^2 \text{ s.t. } \|\mathbf{x}\|_0 \leq T_0. \tag{1}$$

A shape s_t in I_t is represented by a level set function $\Phi_t(\mathbf{u})$. We define $\Phi_t^+(\mathbf{u}) = \Phi_t(\mathbf{u}) + \psi_1$ and $\Phi_t^-(\mathbf{u}) = \Phi_t(\mathbf{u}) - \psi_2$. The regions of interest are two band regions $\Omega_t^1 = \{\mathbf{u} \in \Omega : \Phi_t^-(\mathbf{u}) < 0, \Phi_t(\mathbf{u}) \geq 0\}$, and $\Omega_t^2 = \{\mathbf{u} \in \Omega : \Phi_t^+(\mathbf{u}) > 0, \Phi_t(\mathbf{u}) < 0\}$. We define $\Omega_t^* = \{\mathbf{u} \in \Omega : \Phi_{t-1}^+(\mathbf{u}) + \zeta_1 \geq 0, \Phi_{t-1}^-(\mathbf{u}) - \zeta_2 \leq 0\}$. The constants are chosen such that $s_t \in \Omega_t^*$. Suppose $\{\mathbf{D}_t^1, \mathbf{D}_t^2\}_k$ are two dictionaries adapted to appearance classes Ω_t^1 and Ω_t^2 respectively at scale k. They exclusively span, in terms of sparse representation, the subspaces of the respective classes. Reconstruction residues are defined as

$$\{R_t^c(\mathbf{u})\}_k = \|\mathbf{y}_t^k(\mathbf{u}) - \{\mathbf{D}_t^c \hat{\mathbf{x}}_t^c(\mathbf{u})\}_k\|_2 \tag{2}$$

$\forall \mathbf{u} \in \Omega_t^*$, $k \in \{1, ..., J\}$, and $c \in \{1, 2\}$, where $\hat{\mathbf{x}}_t^c$ is the sparse representation of $\mathbf{y}_t^k$ w.r.t. $\mathbf{D}_t^c$. It is logical to expect that $\{R_t^1(\mathbf{u})\}_k > \{R_t^2(\mathbf{u})\}_k$ when $\mathbf{u} \in \Omega_t^2$, and $\{R_t^1(\mathbf{u})\}_k < \{R_t^2(\mathbf{u})\}_k$ when $\mathbf{u} \in \Omega_t^1$. Combining the multiscale information, we introduce a local appearance discriminant

$$R_t(\mathbf{u}) = \mathbf{1}_{\Omega_t^*}(\mathbf{u}) \sum_{k=1}^{J} (\log \frac{1}{\beta_k}) \text{sgn}(\{R_t^2(\mathbf{u})\}_k - \{R_t^1(\mathbf{u})\}_k), \tag{3}$$

$\forall \mathbf{u} \in \Omega$, where β_k's are the weighting parameters of the J appearance scales.

2.2   Online Multiscale Dictionary Learning

To obtain the discriminant R_t, $\{\mathbf{D}_t^1, \mathbf{D}_t^2\}_k$ and β_k need to be learned. Leveraging the inherent spatio-temporal coherence of individual data, we introduce an online multiscale appearance dictionary learning process supervised in a boosting framework. We interlace the processes of dictionary learning and segmentation as illustrated in Fig. 1. Similar to the database-driven dynamical shape models [3– 5], we also assume a segmented first frame for initialization. It can be achieved by some automatic method with expert correction or purely manual segmentation. We dynamically update the multiscale appearance dictionaries each time a new frame is segmented. For $t > 2$, $\{\mathbf{D}_t^1, \mathbf{D}_t^2\}_k$ are well initialized with $\{\mathbf{D}_{t-1}^1, \mathbf{D}_{t-1}^2\}_k$ and updated with only a few iterations. $\{\mathbf{D}_2^1, \mathbf{D}_2^2\}_k$ are initialized with training samples. To reduce propagation error, we divide a sequence into two subsequences to perform bidirectional segmentation like [8]. The proposed dictionary
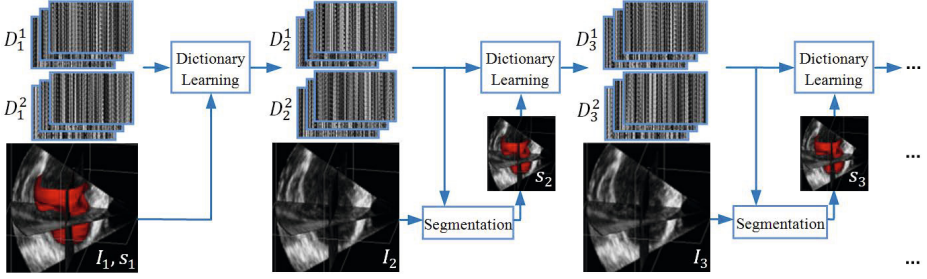
Fig. 1. Dynamical dictionary update interlaced with sequential segmentation

learning algorithm following the structure of the AdaBoost [9] is detailed in Algorithm 1. J dictionary pairs $\{\mathbf{D}_t^1, \mathbf{D}_t^2\}_k$ and weighting parameters β_k are learned from two classes of appearance samples $\{\mathbf{Y}_{t-1}^1\}_k = \{\mathbf{y}_{t-1}^k(\mathbf{u}) : \mathbf{u} \in \Omega_{t-1}^1\}$ and $\{\mathbf{Y}_{t-1}^2\}_k = \{\mathbf{y}_{t-1}^k(\mathbf{u}) : \mathbf{u} \in \Omega_{t-1}^2\}$, $k = 1, ..., J$. The K-SVD [10] algorithm is invoked to enforce the reconstructive property of the dictionaries. The boosting supervision strengthens the discriminative property of the dictionaries and optimizes the weighting of multiscale information.

2.3 MAP Estimation

We estimate the shape Φ_t in frame I_t given the knowledge of $\hat{\Phi}_{1:t-1}$ and $I_{1:t}$. Different from the single-scale method in [8], we integrate a spectrum of complementary multiscale appearance information including intensity, the multiscale local appearance discriminant, and a dynamical shape prediction Φ_t^*. Since Φ_{t-1} and Φ_{t-2} are both spatially and temporally close, we assume a constant evolution speed during $[t - 2, t]$. Within the band domain $\Omega_t^1 \cup \Omega_t^2$ we introduce an approximate shape prediction $\Phi_t^* = \hat{\Phi}_{t-1} + G(\hat{\Phi}_{t-1} - \hat{\Phi}_{t-2})$ to regularize the shape estimation. Here $G(*)$ denotes Gaussian smoothing operation used to preserve the smoothness of level set function. The segmentation is estimated by maximizing the posterior probability:

$$
\begin{aligned}
\hat{\Phi}_t &= \arg\max\nolimits_{\Phi_t} p(\hat{\Phi}_{1:t-1}, I_{1:t-1}, I_t|\Phi_t)p(\Phi_t) \\
&\approx \arg\max\nolimits_{\Phi_t} p(\Phi_t^*, R_t, I_t|\Phi_t)p(\Phi_t) \\
&\approx \arg\max\nolimits_{\Phi_t} p(\Phi_t^*|\Phi_t)p(R_t|\Phi_t)p(I_t|\Phi_t)p(\Phi_t).
\end{aligned}
\tag{4}
$$

The shape regularization is given by $p(\Phi_t^*|\Phi_t)p(\Phi_t) \propto \exp\{-\gamma \int_{\Omega_t^1 \cup \Omega_t^2}(\Phi_t - \Phi_t^*)^2 d\mathbf{u}\} \exp\{-\mu \int_{\Omega} \delta(\Phi_t)|\nabla\Phi_t|d\mathbf{u}\}$. We assume i.i.d. normal distribution of R_t: $p(R_t|\Phi_t) \propto \prod_{\mathbf{u}\in\Omega_t^1} \exp\{\frac{-[R_t(\mathbf{u})-c_1]^2}{2\omega_1^2}\} \prod_{\mathbf{u}\in\Omega_t^2} \exp\{\frac{-[R_t(\mathbf{u})-c_2]^2}{2\omega_2^2}\}$, and i.i.d. Rayleigh density of I_t: $p(I_t|\Phi_t) = \prod_{\mathbf{u}\in\Omega_t^1} \frac{I_t(\mathbf{u})}{\sigma_1^2} \exp\{\frac{-I_t(\mathbf{u})^2}{2\sigma_1^2}\} \prod_{\mathbf{u}\in\Omega_t^2} \frac{I_t(\mathbf{u})}{\sigma_2^2} \exp\{\frac{-I_t(\mathbf{u})^2}{2\sigma_2^2}\}$. Since intensity is not helpful for epicardial discrimination, $p(I_t|\Phi_t)$ is dropped in the epicardial case. The overall segmentation energy functional is given by:

$$
\begin{aligned}
E(\Theta, \Phi_t) &= \int_{\Omega_t^1} I_t^2/2\sigma_1^2 + \log(\sigma_1^2/I_t)d\mathbf{u} + \int_{\Omega_t^2} I_t^2/2\sigma_2^2 + \log(\sigma_2^2/I_t)d\mathbf{u} \\
&+ \int_{\Omega_t^1} (R_t - c_1)^2/2\omega_1^2 d\mathbf{u} + \int_{\Omega_t^2} (R_t - c_2)^2/2\omega_2^2 d\mathbf{u} \\
&+ \gamma \int_{\Omega_t^1 \cup \Omega_t^2} (\Phi_t - \Phi_t^*)^2 d\mathbf{u} + \mu \int_{\Omega} \delta(\Phi_t)|\nabla\Phi_t|d\mathbf{u},
\end{aligned}
\tag{5}
$$

where $\Theta = [c_1, c_2, \omega_1, \omega_2, \sigma_1, \sigma_2]$. We minimize the energy functional as follows: (a) Initialize Φ_t^0 with Φ_{t-1}, $\tau = 0$; (b) Compute the maximum likelihood estimate of $\Theta(\Phi_t^\tau)$; (c) Update $\Phi_t^{\tau+1}$ by gradient descent; (d) Reinitialize $\Phi_t^{\tau+1}$ after every few iterations; (e) Stop if $\|\Phi_t^{\tau+1} - \Phi_t^\tau\|_2 < \xi$, otherwise, $\tau = \tau + 1$, go to (b).

Algorithm 1. Multiscale Appearance Dictionary Learning

Input: appearance samples $\{\mathbf{Y}_{t-1}^1\}_k = \{\mathbf{y}_{1,i}^k\}_{i=1}^{M_1}$ and $\{\mathbf{Y}_{t-1}^2\}_k = \{\mathbf{y}_{2,j}^k\}_{j=1}^{M_2}$, initial dictionaries $\{\mathbf{D}_{t-1}^1, \mathbf{D}_{t-1}^2\}_k$, $k = 1, ..., J$, $\mathbf{w}_1^1 = \{w_{1,i}^1\}_{i=1}^{M_1} = \mathbf{1}$, $\mathbf{w}_2^1 = \{w_{2,j}^1\}_{j=1}^{M_2} = \mathbf{1}$.

Output: dictionary pairs $\{\mathbf{D}_t^1, \mathbf{D}_t^2\}_k$, weighting parameters β_k, $k = 1, ..., J$.

For $k = 1, ..., J$:

- **Resampling:** Draw sample sets $\tilde{\mathbf{Y}}_1^k$ from $\{\mathbf{Y}_{t-1}^1\}_k$ and $\tilde{\mathbf{Y}}_2^k$ from $\{\mathbf{Y}_{t-1}^2\}_k$ based on distributions $\mathbf{p}_1^k = \{p_{1,i}^k\}_{i=1}^{M_1} = \frac{\mathbf{w}_1^k}{\sum_{i=1}^{M_1} w_{1,i}^k}$ and $\mathbf{p}_2^k = \{p_{2,j}^k\}_{j=1}^{M_2} = \frac{\mathbf{w}_2^k}{\sum_{j=1}^{M_2} w_{2,j}^k}$.

- **Dictionary Learning:** Apply the K-SVD to learn $\{\mathbf{D}_t^1, \mathbf{D}_t^2\}_k$ from $\tilde{\mathbf{Y}}_1^k$ and $\tilde{\mathbf{Y}}_2^k$:

$$\min_{\mathbf{D}_t^c, \mathbf{X}} \|\tilde{\mathbf{Y}}_c^k - \mathbf{D}_t^c \mathbf{X}\|_2^2 \quad \text{s.t.} \quad \forall i, \|\mathbf{x}_i\|_0 \le T_0; \ c \in \{1, 2\}. \tag{6}$$

- **Sparse Coding:** $\forall \mathbf{y} \in \{\mathbf{Y}_{t-1}^1, \mathbf{Y}_{t-1}^2\}_k$, solve the sparse representations w.r.t. $\{\mathbf{D}_t^1\}_k$, and $\{\mathbf{D}_t^2\}_k$ using the OMP [11], and get residues $R(\mathbf{y}, \mathbf{D}_t^1)_k$ and $R(\mathbf{y}, \mathbf{D}_t^2)_k$.

- **Classification:** Make a hypothesis $h_k : \mathbf{y} \in \{\mathbf{Y}_{t-1}^1, \mathbf{Y}_{t-1}^2\}_k \rightarrow \{0, 1\}$: $h_k(\mathbf{y}) = Heaviside(R(\mathbf{y}, \mathbf{D}_t^2)_k - R(\mathbf{y}, \mathbf{D}_t^1)_k)$. Calculate the error of h_k: $\epsilon_k = \sum_{i=1}^{M_1} p_{1,i}^k |h_k(\mathbf{y}_{1,i}^k) - 1| + \sum_{j=1}^{M_2} p_{2,j}^k h_k(\mathbf{y}_{2,j}^k)$. Set $\beta_k = \epsilon_k / (1 - \epsilon_k)$.

- **Weight Update:** $w_{1,i}^{k+1} = w_{1,i}^k \beta_k^{1 - |h_k(\mathbf{y}_{1,i}^k) - 1|}$, $w_{2,j}^{k+1} = w_{2,j}^k \beta_k^{1 - h_k(\mathbf{y}_{2,j}^k)}$.

---

3 Experiments and Results

We acquired 26 3D canine echocardiographic sequences from both healthy and post-infarct subjects using a Phillips iE33 ultrasound imaging system with a frame rate of ~ 40 Hz. Each sequence spanned a cardiac cycle. The sequential segmentation was initialized with a manual tracing of the first frame. Both endocardial and epicardial borders were segmented throughout the sequences. Fig. 2 shows typical segmentation examples by our method. 100 frames were randomly drawn from ~ 700 frames for manual segmentation and quality assessment. We evaluated automatic results against expert manual tracings using the following segmentation quality metrics: Hausdorff Distance (HD), Mean Absolute Distance (MAD), Dice coefficient (DICE), and Percentage of True Positives (PTP).

Benefit from the Dynamical Appearance Model. When the dynamical appearance components are turned off, our model reduces to a conventional ultrasound intensity model: the Rayleigh model [12]. Comparison with the Rayleigh method clearly shows the added value of the proposed DAM. Since the Rayleigh method is generally sensitive to initial contours, we initialized its segmentation of each frame with the first frame manual tracing. Fig. 3(a) compares typical
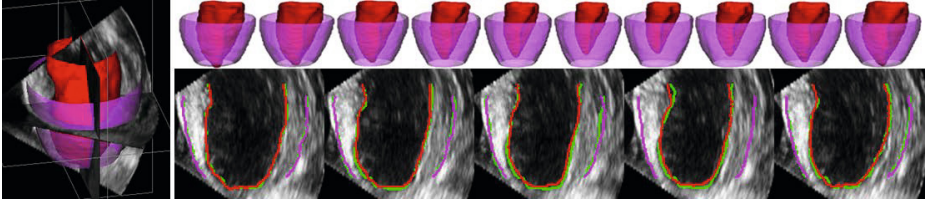
Fig. 2. Typical segmentations by our method (red,purple) and manual tracings (green)

segmentation examples by the Rayleigh method and our method. We observed that the Rayleigh method was easily trapped by misleading intensity information (e.g., image inhomogeneities and artifacts), while our approach produced accurate segmentations. Fig. 2 qualitatively shows the capability of the DAM in estimating reliably 3D left ventricular borders throughout the whole cardiac cycle. The Rayleigh method did not generate acceptable segmentation sequences in the experiment. Table 1 demonstrates that the DAM significantly outperformed the Rayleigh model. Better means (higher DICE and PTP, and lower MAD and HD) and lower standard deviations show the remarkable improvement of segmentation accuracy and robustness achieved by employing the DAM.

Table 1. Sample means and standard deviations of segmentation quality measures

expressed as mean±std		DICE (%)	PTP (%)	MAD (mm)	HD (mm)
Endocardial	Rayleigh [12]	74.9 ± 18.8	83.1 ± 16.3	2.01 ± 1.22	9.17 ± 3.37
	DAM	93.6 ± 2.49	94.9 ± 2.34	0.57 ± 0.14	2.95 ± 0.62
	SSDM [5]	——	95.9 ± 1.24	1.41 ± 0.40	2.53 ± 0.75
Epicardial	Rayleigh [12]	74.1 ± 17.4	82.5 ± 12.0	2.80 ± 1.55	16.9 ± 9.30
	DAM	97.1 ± 0.93	97.6 ± 0.86	0.60 ± 0.19	3.03 ± 0.76
	SSDM [5]	——	94.5 ± 1.74	1.74 ± 0.39	2.79 ± 0.97

Advantages over Single-scale Sparse Representation. We compared our DAM to the single-scale sparse representation model (SSR) in [8]. The SSR was extended to 3D and performed at 5 appearance scales ranging from low scale $3.5 \times 3.5 \times 3.5 mm^3$ to high scale $15.5 \times 15.5 \times 15.5 mm^3$, while the DAM utilized
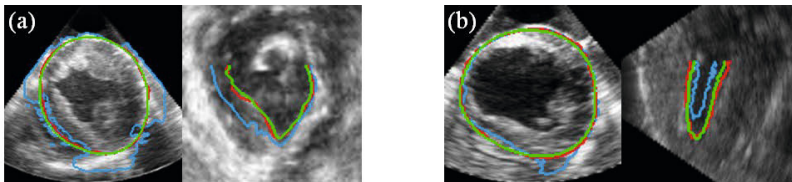
Fig. 3. Segmentation examples. (a) Manual (Green), DAM (Red), Rayleigh (Blue). (b) Manual (Green), DAM (Red), SSR (Blue).

multiscale appearance information. Fig. 3(b) presents end-systolic segmentation examples showing that the use of DAM resulted in lower propagation error and higher segmentation accuracy compared to the SSR. Fig. 4 presents the quantitative results of the comparison study. We observed that the performance of the SSR varied with the scale, which implies its sensitivity to the appearance scale. The SSR required scale tuning to get better results. The DAM achieved the best results in almost all the metrics for both endocardial and epicardial segmentations, which demonstrates the advantages of DAM over SSR. By summarizing complementary multiscale appearance information, the DAM consistently produced accurate segmentations without careful parameter tuning.
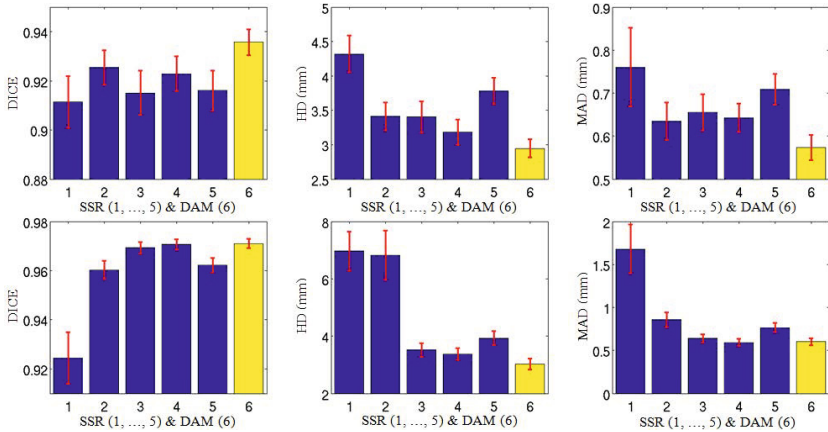
Fig. 4. Means and 95% confidence intervals obtained by the SSR (blue, scales 1, ..., 5) and the DAM (yellow, 6) in endocardial (top row) and epicardial (bottom row) cases.

Comparison with Database-Driven Dynamical Models. Table 1 compares the HD, MAD and PTP achieved by our model and that by a state-of-the-art database-driven dynamical shape model SSDM reported in [5]. The database-free DAM achieved comparable results with the SSDM, and outperformed the SSDM in segmenting epicardial borders. It is worth noticing that the DAM does not require more human interaction at the segmentation stage than the database-driven dynamical models such as [3–5] which also need manual tracings of the first or first few frames for initialization.

4   Discussion and Conclusion

We have proposed a 3D dynamical appearance model that exploits the inherent spatio-temporal coherence of individual echocardiographic data. It employs multiscale sparse representation, online multiscale appearance dictionary learning, and a spectrum of complementary multiscale appearance information including intensity, multiscale local appearance, and shape. Our method resulted in

significantly improved accuracy and robustness of left ventricular segmentation compared to a standard intensity method and our previous single-scale sparse representation method. The DAM achieved comparable results with a state-of-the-art database-driven statistical dynamical model SSDM. Since the DAM is database-free, it overcomes the limitations introduced by the use of databases. The DAM can be applied to the cases (e.g., the post-infarct subjects in this study) where it is inappropriate to apply database-based a priori motion or shape knowledge. Even when the priors are effective, the DAM can be a good choice for complementing the database and relaxing the reliance of statistical models (e.g., [2–6]) on database quality. Future work includes extensions to human data, other modalities, and an integrated online and offline learning framework to exploit their complementarity.

References

1. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active appearance models. IEEE TPAMI 23(6), 681–685 (2001)
2. Bosch, J.G., Mitchell, S.C., Lelieveldt, B.P.F., Nijland, F., Kamp, O., Sonka, M., Reiber, J.H.C.: Automatic segmentation of echocardiographic sequences by active appearance motion models. IEEE TMI 21(11), 1374–1383 (2002)
3. Jacob, G., Noble, J.A., Behrenbruch, C.P., Kelion, A.D., Banning, A.P.: A shape-space based approach to tracking myocardial borders and quantifying regional left ventricular function applied in echocardiography. IEEE TMI 21(3), 226–238 (2002)
4. Sun, W., Çetin, M., Chan, R., Reddy, V., Holmvang, G., Chandar, V., Willsky, A.S.: Segmenting and Tracking 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)
5. Zhu, Y., Papademetris, X., Sinusas, A.J., Duncan, J.S.: A Dynamical Shape Prior for LV Segmentation from RT3D Echocardiography. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009, Part I. LNCS, vol. 5761, pp. 206–213. Springer, Heidelberg (2009)
6. Yang, L., Georgescu, B., Zheng, Y., Meer, P., Comaniciu, D.: 3D ultrasound tracking of the left ventricle using one-step forward prediction and data fusion of collaborative trackers. In: CVPR (2008)
7. Noble, J.A., Boukerroui, D.: Ultrasound image segmentation: a survey. IEEE TMI 25(8), 987–1010 (2006)
8. Huang, X., Lin, B.A., Compas, C.B., Sinusas, A.J., Staib, L.H., Duncan, J.S.: Segmentation of left ventricles from echocardiographic sequences via sparse appearance representation. In: MMBIA, pp. 305–312 (January 2012)
9. Freund, Y., Schapire, R.: A Desicion-Theoretic Generalization of On-line Learning and an Application to Boosting. In: Vitányi, P.M.B. (ed.) EuroCOLT 1995. LNCS, vol. 904, pp. 23–37. Springer, Heidelberg (1995)
10. Aharon, M., Elad, M., Bruckstein, A.: K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. IEEE TSP 54(11), 4311–4322 (2006)
11. Tropp, J.: Greed is good: algorithmic results for sparse approximation. IEEE TIT 50(10), 2231–2242 (2004)
12. Sarti, A., Corsi, C., Mazzini, E., Lamberti, C.: Maximum likelihood segmentation of ultrasound images with rayleigh distribution. IEEE TUFFC 52(6), 947–960 (2005)

Automatic Detection and Segmentation of Kidneys in 3D CT Images Using Random Forests

Rémi Cuingnet[1], Raphael Prevost[1,2], David Lesage[1], Laurent D. Cohen[2], Benoît Mory[1], and Roberto Ardon[1]

[1] Philips Research Medisys, France
[2] CEREMADE, UMR 7534 CNRS, Paris Dauphine University, France

Abstract. Kidney segmentation in 3D CT images allows extracting useful information for nephrologists. For practical use in clinical routine, such an algorithm should be fast, automatic and robust to contrast-agent enhancement and fields of view. By combining and refining state-of-the-art techniques (random forests and template deformation), we demonstrate the possibility of building an algorithm that meets these requirements. Kidneys are localized with random forests following a coarse-to-fine strategy. Their initial positions detected with global contextual information are refined with a cascade of local regression forests. A classification forest is then used to obtain a probabilistic segmentation of both kidneys. The final segmentation is performed with an implicit template deformation algorithm driven by these kidney probability maps. Our method has been validated on a highly heterogeneous database of 233 CT scans from 89 patients. 80 % of the kidneys were accurately detected and segmented (Dice coefficient > 0.90) in a few seconds per volume.

1 Introduction

Segmentation of medical images is a key step to gathering anatomical information for diagnosis or interventional planning. Renal volume and perfusion, which can be extracted from CT images, are typical examples for nephrologists. However, it is often long and tedious for clinicians to segment 3D images. Automatic and fast segmentation algorithms are thus needed for practical use. It is yet still challenging to design an algorithm robust enough to noise, acquisition artifacts or leakages in neighboring organs.

Several papers in the literature tackle the problem of kidney segmentation in CT images. In [1] and [2], the authors used the Active Shape Model framework to learn the kidney mean shape and principal modes of variation, in order to constrain the segmentation. Recently Khalifa et al. [3] proposed a level-set approach, based on a new force combining shape and intensity priors as well as spatial interactions, which showed promising results. However, they were assessed on small datasets (41, 17 and 20 volumes in [1], [2] and [3] respectively). Moreover, all these algorithms are either based on a manual initialization, or tested

on images already cropped around the kidney. A fully automatic method has already been introduced by Tsagaan et al. [4], but their detection of the region of interest presents limitations. First, it relies on hard geometrical constraints, which requires knowledge on the field of view. Then, a rough search is done by template matching, which is not robust to pathologies or kidney orientation.

In this paper, we propose a fast and completely automatic method to detect and segment both kidneys in any kind of CT image: acquired at different contrast phases (or without contrast) with various fields of view, from both healthy subjects and patients with kidney tumors. Kidneys' positions are first detected with regression forests following a coarse to fine strategy (Section 2). Then a two-step segmentation is performed on cropped images around the kidneys (Section 3) using (i) a random forest to estimate a probability map of each kidney and (ii) a template deformation algorithm [5] to extract the kidney surface. Experiments and results are detailed in Section 4.

2   Kidney Detection with Regression Forests

This section presents a fast and reliable estimation of the kidneys' locations. Various approaches for anatomy detection and localization have been proposed in the literature (Section 2.1). We propose a regression-based method in two steps. The whole image is first used to provide an estimate of the region of interest (Section 2.2) which is then refined using local information only (Section 2.3).

2.1   Background on Organ Detection

Registration-based approaches using labeled atlases (e.g. [6]) have often been used for this problem. However such approaches are subject to registration errors due to inter individual variability. The robustness of the registration step can be improved by using multi-atlas or multi-template techniques [7] but at the cost of an increase in computation time.

More recently, supervised learning methods have been used for this detection problem to better take into account interindividual variability. Most ***classification-based*** detection algorithms consist in constructing a classifier whose role is to predict from local features to which organ a voxel belongs (e.g. [8]). However, by considering only local features, such approaches do not benefit from anatomical contextual information. To overcome this shortcoming, Criminisi et al. [9] used a generalization of Haar features that models contextual information. Instead of classifying each voxel, some authors consider the detection problem as finding a vector of parameters describing the organ locations. Such parameters can describe contour line or surface of an organ [10] or more simply bounding boxes around the different organs of interest [11]. The role of the classifier is then to predict whether a set of parameters is correct or not. Zheng et al. [11] used a greedy approach to avoid searching the whole parameter space, which is intractable.

Zhou et al. [12] showed that finding a set of continuous parameters from an image is by definition a multiple-output regression problem. More precisely, they

proposed a boosting ridge regression to detect and localize the left ventricle in cardiac ultrasound 2D images. ***Regression-based*** techniques do not require an exhaustive search of parameters. Other regressors such as regression forests and random ferns have also been proposed [13,14].

In the following, we consider regression forests to simultaneously detect both kidneys. Regression forests [15,16] are particularly well adapted to this problem in clinical routine since, thanks to their tree structures, they allow very fast testing with nonlinear regression functions. Since there is no explicit regularization, random forests require a large number of training samples to avoid overfitting the training data. Here, this is not a limitation since the training samples are the voxels of the training CT scans.

2.2   Coarse Localization of the Kidneys

We consider the detection step as the problem of finding bounding boxes around both kidneys. First, we find a coarse positioning based on contextual information adapting the approach proposed by Criminisi et al. [13]. Then, the position of each box is refined based on local information.

Each bounding box is parameterized by a vector in $\mathbb{R}^6$ (two points in 3D). A random forest is trained on CT scans with known kidney bounding boxes to predict for each voxel the relative position and size of the kidneys. Since CT intensities (expressed in Hounsfield units) have direct physical meaning, as explained in [13], the features used are the mean intensities over displaced, asymmetric cuboidal regions. To allow a much faster training, we used residual sum of squares (RSS) instead of the information gain for the node optimization in the training stage. Note that optimizing the RSS comes to minimizing the trace of the covariance matrix at each node instead of its determinant. We did not notice any differences in term of prediction accuracies.

This step gives a first estimate of the kidneys' positions and sizes. By construction, the relative estimated position of the left and right kidneys are strongly correlated. Such a correlation ensures coherent results but may not reflect the whole possible interindividual variability. This might be critical when the number of subjects in the training set is low. To overcome this shortcoming, we propose a refinement step of the bounding boxes that relaxes the correlation between the two kidneys' position.

2.3   Refinement of the Region of Interest

This step consists in refining the left and right kidneys' positions based on local information only. The constraints between the kidneys' relative positions are relaxed by treating both kidneys independently. For each kidney, a regression forest is trained to predict, from every voxel located in its neighborhood, the relative position of the kidney's center.We used the same training set as in the previous step. The features used for this step are, for each voxel, its intensity and its gradient magnitude, as well its neighbors'.

For testing, only the voxels in the neighborhood of the center of the bounding box predicted by the first step are considered. As depicted in Figure 1.b, each voxel v then votes for a location $\hat{\mathbf{c}}_v$ of the kidney's center. For robustness sake, the final location estimate is $\hat{\mathbf{c}} = \text{argmin}_{\mathbf{c} \in \mathbb{R}^3} \sum_{v=1}^{K} \|\mathbf{c} - \hat{\mathbf{c}}_v\|_1$ where $(\hat{\mathbf{c}}_v)_{1,K}$ are the K votes with the highest probability. The final bounding box is then translated accordingly.

To ensure stability, this refinement step is constrained to very small displacements and is iterated until convergence. This can be considered as a cascaded pose regression similar to [17]. Illustration of the kidney detection is given in Figure 1 and quantitative results are reported in Section 4.

(a)   (b)   (c)
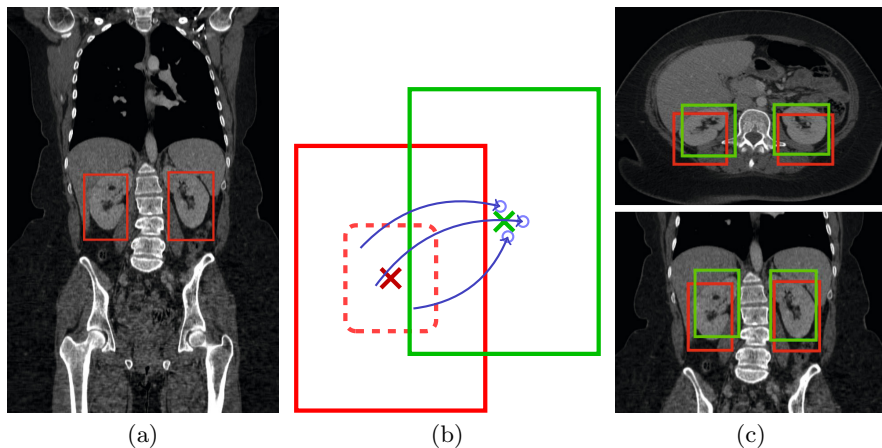
Fig. 1. Illustration of the kidney detection on a CT volume. (a) Initial bounding boxes detected using global contextual information. (b) Refinement step: voxels near the center of the initial bounding box (red) vote for its new center, using only local information. (c) Comparison between the initial (red) and refined (green) bounding box.

3   Kidney Segmentation

Even when the image is cropped to a region Ω around the kidney, its segmentation remains a challenging task: (i) kidneys are composed of different tissues (cortex, medulla, sinus) resulting in different image intensities, (ii) surrounding organs may touch the kidney without a clear boundary, (iii) the contrast phase of the CT image is unknown. For all these reasons, it is not possible to solely rely on the image intensity, and we rather use it simultaneously with other features.

3.1   Probability Estimation via Random Forests

In addition to regression, random forests can also be used to perform classification [15,16]. We trained a random forest classifier to predict, for each voxel

x of the previously detected bounding box, the probability $P(\mathbf{x})$ of belonging to a kidney. This random forest combines different image features: intensity and first/second order derivatives of the voxel and its neighbors. Decision stumps were used as weak classifiers and the impurity criterion was the Gini index [15,16]. Such probability maps are shown in Figures 2.a and 2.c. Independently from the contrast-phase, the whole kidney tissues are enhanced, whereas the confusing adjacent structures are removed.

3.2 Initialization of the Segmentation

For the sake of robustness to interindividual variability and to pathologies, we only assumed that kidneys have a bean shape that can be roughly approximated by an ellipsoid. The segmentation algorithm is thus initialized with the ellipsoid $\mathcal{E} = \{\mathbf{x} \in \mathbb{R}^3 \mid (\mathbf{x}-\mathbf{c}_{\mathcal{E}})^T \mathcal{M}_{\mathcal{E}}^{-1}(\mathbf{x}-\mathbf{c}_{\mathcal{E}}) = 1\}$, where $\mathbf{c}_{\mathcal{E}} = \int_{\Omega} P(\mathbf{x}) \, \mathbf{x} \, d\mathbf{x}$ denotes the weighted barycenter and $\mathcal{M}_{\mathcal{E}}$ is proportional to the weighted covariance matrix $\int_{\Omega} P(\mathbf{x}) \, (\mathbf{x} - \mathbf{c}_{\mathcal{E}})(\mathbf{x} - \mathbf{c}_{\mathcal{E}})^T \, d\mathbf{x}$.

3.3 Implicit Template Deformation

We followed the framework introduced in [5] to deform the ellipsoid $\mathcal{E}$. A model-based approach is here particularly suited because (i) kidneys usually have very smooth shapes, (ii) we want the algorithm to reasonably extrapolate the boundary when the probability map is uncertain. Hereafter we recall the main principles of the adapted model-based deformation algorithm.

Given a working image $I : \Omega \to \mathbb{R}$ and the initial ellipsoid $\mathcal{E}$ defined by an implicit function ϕ, we find a transformation $\psi : \Omega \to \Omega$ such that the image gradient flux across the surface of the deformed ellipsoid $\mathcal{E}(\psi) = (\phi \circ \psi)^{-1}(0)$ is maximum. Denoting **n** the normal vector, the segmentation energy is then

$$E_s(\psi) = \int_{\mathcal{E}(\psi)} - \langle \boldsymbol{\nabla} \boldsymbol{I}(\mathbf{x}) \, , \, \mathbf{n}(\mathbf{x}) \rangle \; d\mathbf{x} + \lambda \mathcal{R}(\psi) \; , \tag{1}$$

$\mathcal{R}(\psi)$ is a regularization term which prevents large deviations from the original ellipsoid. The transformation is decomposed as $\psi = \mathcal{L} \circ \mathcal{G}$ where

- $\mathcal{G}$ is a global linear transformation, which may correct or adjust the center, orientation and scales of the initial ellipsoid;
- $\mathcal{L}$ is a non-rigid local deformation, expressed using a displacement field **u** such that $\mathcal{L}(\mathbf{x}) = \mathbf{x} + (\mathbf{u} * K_\sigma)(\mathbf{x})$. K_σ is a Gaussian kernel that provides built-in smoothness, at a given scale σ.

This decomposition allows $\mathcal{R}$ to be pose-invariant and constrains only the non-rigid deformation : $\mathcal{R}(\psi) = \mathcal{R}(\mathcal{L}) = \int_{\Omega} \|\mathcal{L} - Id\|^2 = \int_{\Omega} \|\mathbf{u} * K_\sigma\|^2$. Finally, using Stokes formula, E_s can be rewritten as

$$E_s(\psi) = - \int_{\Omega} H(\phi \circ \mathcal{L} \circ \mathcal{G}) \; \Delta I + \lambda \int_{\Omega} \|\mathbf{u} * K_\sigma\|^2 \; , \tag{2}$$

where H is the Heaviside function and Δ is the Laplacian operator. This energy is minimized, with respect to the parameters of $\mathcal{G}$ and each component of the vector field $\mathbf{u}$, through a gradient descent. Note that since the energy in (2) is not convex, the resulting segmentation depends on the initialization. Hence, we first apply this algorithm to the probability map ($I = P$) in order to reach an appropriate local minimum (e.g. no leaks in surrounding tissues). The segmentation is finally refined on the original CT volume, with a higher shape constraint parameter λ and a finer scale σ (Figures 2.b and 2.d).

(a)   (b)   (c)   (d)
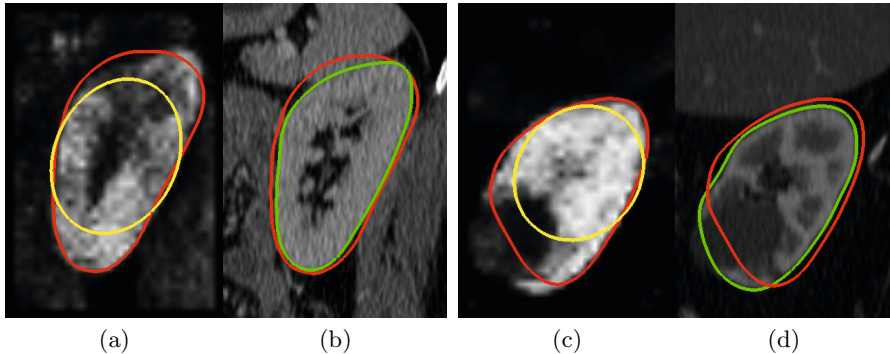
Fig. 2. Illustration of the two-step kidney segmentation on two cases: (a-b) non-contrasted volume of a healthy patient, (c-d) contrast-enhanced image of a kidney with a tumor. The kidney probability maps (a) and (c) are learned with a random forest, and used to coarsely segment the kidney (red) by deforming an initial ellipsoid (yellow). The segmentation is then refined (green) using the original volumes (b) and (d).

4   Experiments and Results

The validation of our method was performed on a representative clinical dataset of 233 CT volumes from 89 subjects including diseased patients. The scans were contrast-enhanced or not and with various fields of view and spatial resolutions. They have between 33 and 973 (mean: 260) 512×512 slices with slice (resp. interslice) resolutions ranging from 0.5 to 1 mm (resp. 0.5 to 3.0 mm). 16% of the kidneys were slighlty truncated, but were nevertheless included in the evaluation to keep it clinically representative. The database was split into a training set of 54 volumes from 26 randomly selected patients, and a testing set composed of the other 179 volumes from 63 patients.

The proposed algorithm used 3 regression forests and 2 classification forests. Each forest was composed of 7 trees with a maximum tree depth $d = 15$ and a minimal node size $n = 100$. We did not notice a high sensitivity of the results to these parameters value. The whole training procedure lasts ~ 5 hours. Times are indicated for a C++ implementation (3.0 GHz dual-core, 4 Go RAM).

Kidney Detection. Detection errors were defined as the absolute difference between predicted and true wall positions averaged over all the bounding box sides. The distance between the predicted bounding box center and the ground truth was also used to assess the detection accuracy. These results are given in Table 1 and compared to those reported in [13]. The refinement step (Section 2.3), for a low extra time cost, greatly increases the accuracy of the bounding box detection (e.g. the median center error is divided by 3).

Table 1. Detection results reported as: Mean ± Standard-deviation (Median)

Detection	Walls error (mm)		Center error (mm)		Time (s)
	Left	Right	Left	Right	Left+Right
[13]	17 ± 17 (13)	19 ± 18 (12)	–	–	–
Coarse	12 ± 7 (10)	13 ± 6 (11)	23 ± 14 (20)	26 ± 13 (23)	2.1 ± 0.5 (2.0)
Refined	7 ± 10 (5)	7 ± 6 (6)	11 ± 18 (6)	10 ± 12 (7)	2.8 ± 1.7 (2.4)

Automatic Segmentation. The results of the automatic segmentation including the detection step were compared to the ground truth using the Dice index. Figure 3 shows the histograms of the scores for both kidneys. 80 % of the kidneys were correctly detected and segmented (Dice > 0.90). The algorithm failed in only 6% of the cases (Dice < 0.65). The total execution time is around 10 s.

	Dice	Left	Right
1st **quartile**		0.93	0.93
median		0.96	0.96
3rd **quartile**		0.97	0.97
maximum		0.99	0.99
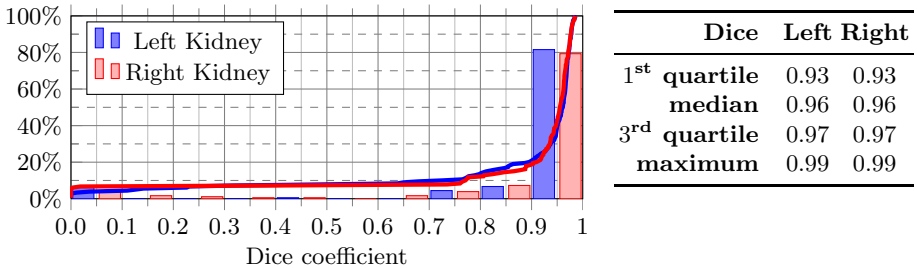
Fig. 3. Distribution of the Dice coefficient between the ground truth and the automatically segmented kidneys. Red and blue lines show the cumulative distribution.

5   Conclusion

This paper presented a fully automatic method to detect and segment both kidneys in any CT volume using random regression and classification forests. Regression forests were used to estimate the kidneys' positions. A classification forest was then used to obtain a probability map of each kidney. The segmentation was carried out with an implicit template deformation algorithm. The

full automation and the execution time are compatible with clinical routine. Results show that our method provides an accurate segmentation in 80% of the cases despite the highly heterogeneous database. Remaining cases were mostly due to pathological kidneys not represented in the training set. Such cases could be quickly corrected by the clinician, since the chosen model-based deformation algorithm [5] allows user interactions. We also emphasize the generality of our framework, that could be as future work extended to other organs.

References

1. 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)
2. Li, X., Chen, X., Yao, J., Zhang, X., Tian, J.: Renal Cortex Segmentation Using Optimal Surface Search with Novel Graph Construction. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part III. LNCS, vol. 6893, pp. 387–394. Springer, Heidelberg (2011)
3. Khalifa, F., Elnakib, A., Beache, G.M., Gimel'farb, G., El-Ghar, M.A., Ouseph, R., Sokhadze, G., Manning, S., McClure, P., El-Baz, A.: 3D Kidney Segmentation from CT Images Using a Level Set Approach Guided by a Novel Stochastic Speed Function. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part III. LNCS, vol. 6893, pp. 587–594. Springer, Heidelberg (2011)
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, Part I. LNCS, vol. 2488, pp. 556–563. Springer, Heidelberg (2002)
5. Mory, B., Somphone, O., Prevost, R., Ardon, R.: Real-Time 3D Image Segmentation by User-Constrained Template Deformation. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part I. LNCS, vol. 7510, pp. 560–567. Springer, Heidelberg (2012)
6. Fenchel, M., Thesen, S., Schilling, A.: Automatic Labeling of Anatomical Structures in MR FastView Images Using a Statistical Atlas. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 576–584. Springer, Heidelberg (2008)
7. Isgum, I., et al.: Multi-atlas-based segmentation with local decision fusion: Application to cardiac and aortic segmentation in CT scans. IEEE Trans. Med. Imaging 28(7), 1000–1010 (2009)
8. Montillo, A., Shotton, J., Winn, J., Iglesias, J.E., Metaxas, D., Criminisi, A.: Entangled Decision Forests and Their Application for Semantic Segmentation of CT Images. In: Székely, G., Hahn, H.K. (eds.) IPMI 2011. LNCS, vol. 6801, pp. 184–196. Springer, Heidelberg (2011)
9. Criminisi, A., et al.: Decision Forests with Long-Range Spatial Context for Organ Localization in CT Volumes. In: MICCAI Workshop PMMIA (2009)
10. Georgescu, B., et al.: Database-guided segmentation of anatomical structures with complex appearance. In: CVPR, vol. 2, pp. 429–436. IEEE (2005)
11. Zheng, Y., et al.: Four-chamber heart modeling and automatic segmentation for 3-D cardiac CT volumes using marginal space learning and steerable features. IEEE Trans. Med. Imaging 27(11), 1668–1681 (2008)

12. Zhou, S., et al.: Image based regression using boosting method. In: ICCV, vol. 1, pp. 541–548. IEEE (2005)
13. Criminisi, A., Shotton, J., Robertson, D., Konukoglu, E.: Regression Forests for Efficient Anatomy Detection and Localization in CT Studies. In: Menze, B., Langs, G., Tu, Z., Criminisi, A. (eds.) MICCAI 2010. LNCS, vol. 6533, pp. 106–117. Springer, Heidelberg (2011)
14. Pauly, O., Glocker, B., Criminisi, A., Mateus, D., Möller, A.M., Nekolla, S., Navab, N.: Fast Multiple Organ Detection and Localization in Whole-Body MR Dixon Sequences. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part III. LNCS, vol. 6893, pp. 239–247. Springer, Heidelberg (2011)
15. Breiman, L.: Random forests. Machine Learning 45(1), 5–32 (2001)
16. Criminisi, A., et al.: Decision forests for classification, regression, density estimation, manifold learning and semi-supervised learning. Technical report, Microsoft Research (2011)
17. Dollar, P., et al.: Cascaded pose regression. In: CVPR, pp. 1078–1085. IEEE (2010)

Neighbourhood Approximation Forests

Ender Konukoglu, Ben Glocker, Darko Zikic, and Antonio Criminisi

Microsoft Research Cambridge

Abstract. Methods that leverage neighbourhood structures in high-dimensional image spaces have recently attracted attention. These approaches extract information from a new image using its "neighbours" in the image space equipped with an application-specific distance. Finding the neighbourhood of a given image is challenging due to large dataset sizes and costly distance evaluations. Furthermore, automatic neighbourhood search for a new image is currently not possible when the distance is based on ground truth annotations. In this article we present a general and efficient solution to these problems. "Neighbourhood Approximation Forests" (NAF) is a supervised learning algorithm that approximates the neighbourhood structure resulting from an arbitrary distance. As NAF uses only image intensities to infer neighbours it can also be applied to distances based on ground truth annotations. We demonstrate NAF in two scenarios: i) choosing neighbours with respect to a deformation-based distance, and ii) age prediction from brain MRI. The experiments show NAF's approximation quality, computational advantages and use in different contexts.

1 Introduction

Computational methods that leverage available datasets for analyzing new images show high accuracy and robustness. Among these methods one class that has lately shown significant potential is *neighbourhood*-based approaches. These approaches formulate the set of all images as a high-dimensional space equipped with an application-specific distance. They then utilize the neighbourhood structure of this space for various tasks. The underlying principle is that neighbouring images, in other words images that are similar with respect to the distance, provide valuable and accurate information about each other. Therefore, when analyzing a new image one can propagate information from its neighbours.

Neighbourhood-based approach, as a general framework, has recently been applied in different contexts. Patch-based techniques [7] and multi-atlas based methods [11] utilize it for segmenting medical images. Nonlinear "manifold"-based methods, which are used in different applications [10,20], also rely on the neighbourhood-based approach, i.e. the neighbourhood structure is preserved during the low-dimensional embedding and subsequent analyses in the low-dimensional space are based on this structure.

One problem in neighbourhood-based approaches, which currently limits their use, is determining the close neighbours of a new image within an existing

dataset. In theory, to determine this neighbourhood one should compute the distances between the new image and all the other images. However, depending on the nature of the distance and the size of the training set this exhaustive search can be computationally very expensive or even impossible. For instance, in multi-atlas based segmentation one would register a new image to all other images to determine its neighbours and propagate labels based on this. The cost of this exhaustive search is high due to computational times of nonlinear registration. Similar problems exist in "manifold"-based techniques, as also pointed out in [2]. In case the distance is defined with respect to ground truth annotation not available for the new image then exhaustive search becomes impossible.

Besides exhaustive search, currently used techniques for finding the neighbourhood of a new image is either through heuristic search strategies [1,7] or K-means like approaches such as multi-template constructions [16,3]. Heuristic strategies are based on application specific rules, therefore not flexible. K-means like approaches have a trade-off in choosing the number of centroids, i.e. too many will result in a computational bottleneck and too few will not correctly reflect the neighbourhood structure. In manifold techniques, some methods find the manifold coordinates of a new image without reconstructing the embedding, [5,14]. However, these methods also rely on computing all distances. Lastly, if a set of low-dimensional features that describes the neighbourhood structure is known then quantization [13] and hashing [19,15] techniques create short binary codes from these features for fast image retrieval. Construction of the initial low-dimensional features still remains an open problem though.

In this article, we present "Neighborhood Approximation Forests" (NAF), a general supervised learning algorithm for approximating an arbitrary neighbourhood structure using image intensities. The main principle of NAF is to learn a compact representation that can describe the neighbourhood structure of a high-dimensional image space equipped with a user-specified distance. For a new image, NAF predicts its neighbourhood within an existing dataset in an efficient manner. We first define the general framework of neighbourhood-based approaches and detail the proposed algorithm. In the experiments we apply NAF to two applications. First we treat the problem of determining the closest neighbours of a new image within a training set with respect to the amount of deformation between images. This experiment demonstrates the prediction accuracy of NAF compared to the real neighbourhood structure and shows the computational advantages. In the second application we devise a simple neighbourhood-based regression method powered by NAF to solve the "toy" problem of age prediction using brain MRI. This experiment demonstrates the use of NAF on an image space where the neighbourhood relation is determined by a continuous and non-image based meta information. Results show high regression accuracies achieved by NAF compared to the values reported in the literature.

2   Neighbourhood Approximation Forests

Neighbourhood-based approach (NbA) is a general framework that is applied for various image analysis tasks. The underlying principle is to extract information

from an image using other "similar" images within a dataset with ground truth, i.e. *training set*. NbA formulates the set of all images as a high-dimensional space $\mathcal{I}$, where each point $I \in \mathcal{I}$ denotes an image. The dataset with ground truth is a finite subset within this space $\mathbf{I} = \{I_p\}_{p=1}^{P} \in \mathcal{I}$. The space $\mathcal{I}$ is equipped with a distance $\rho(I, J)$ that quantifies a similarity between images, which is application dependent. For an image I the set of k most similar images in $\mathbf{I}$ is then defined as the neighborhood $\mathbf{N}_{\rho}^{k}(I)$, i.e. k images with the lowest distance to I. To analyse a new image $J \notin \mathbf{I}$, one needs to determine $\mathbf{N}_{\rho}^{k}(J)$ within $\mathbf{I}$ to be able to use NbA. This is challenging because the computation of $\rho(\cdot, \cdot)$ between J and all images in $\mathbf{I}$ can be expensive or even not possible. In the following we describe a learning algorithm to approximate $\mathbf{N}_{\rho}^{k}(J)$ that overcomes these challenges.

Our approach relies on the hypothesis that the neighbourhood structure constructed by $\rho(\cdot, \cdot)$ can be approximated using compact image descriptors derived from intensity information. Consequently, using these descriptors, for a new image J we can approximate its neighborhood $\mathbf{N}_{\rho}^{k}(J)$ within $\mathbf{I}$ without the need to evaluate $\rho(\cdot, \cdot)$. Neighborhood Approximation Forests (NAF) is a supervised algorithm that learns such descriptors for arbitrary $\rho(\cdot, \cdot)$. It is a variant of random decision forests [6,8], i.e. an ensemble of binary decision trees, where each tree is an independently learned predictor of $\mathbf{N}_{\rho}^{k}(J)$ given J. As all supervised learning algorithms NAF has two phases: training and prediction. Below we explain these phases and then demonstrate NAF in Section 3.

Predicting neighbourhood with a single tree: We represent each image I using a set of intensity-based features $\mathbf{f}(I) \in \mathbb{R}^Q$ of possibly high dimensions, which can be as simple as intensity values at different points. These features have no prior-information on $\rho(\cdot, \cdot)$. For a new image J, each tree T predicts J's neighbours within a training set $\mathbf{I}$ by applying a sequence of learned binary tests to a subset of its entire feature vector $\mathbf{f}_T(J) \in \mathbb{R}^q$, $q < Q$ and $\mathbf{f}_T(J) \subset \mathbf{f}(J)$. Each binary test in the sequence depends on the result of the previous test. This whole process is represented as a binary decision tree [4], where each test corresponds to a branching node in the tree. Starting from the root node s_0 the image J traverses the tree taking a specific path and arrives at a node with no further children, a leaf-node. The path and the final leaf-node depend on the feature vector $\mathbf{f}_T(J)$ and the binary tests at each node.

Each leaf-node stores the training images (or simply their indeces) $I_n \in \mathbf{I}$ which traversed T and arrived at that node. So, at the leaf-node J arrives there is a subset of training images which have taken the same path as J and therefore share similar feature values based on the applied tests. This subset of training images, $\mathbf{N}_{T(\rho)}(J)$, is the neighbourhood of J predicted by T. The subscript $T(\rho)$ denotes the tree's dependence on $\rho(\cdot, \cdot)$, which we explain in the training part.

Approximating neighborhood with the forest: The forest F is composed of multiple independent trees with independent predictions. Each tree works with a different subset of features $\mathbf{f}_T(J) \subset \mathbf{f}(J)$ focusing on a different part of the feature space. We compute the ensemble forest prediction by combining the independent tree predictions. This combination process computes the approximate affinity of J to each I_n by $\mathbf{w}_F(J, I_n) \triangleq \sum_{\forall T \in F} \mathbf{1}_{\mathbf{N}_{T(\rho)}(J)}(I_n)$, where $\mathbf{1}_A(x)$

is the indicator function (we note that [9] uses a similar construction for a different purpose: defining a neighborhood structure) The forest prediction of $\mathbf{N}_\rho^k(J)$ is simply the k training images with the largest $\mathbf{w}_F(J, I_n)$ values. We denote this set with $\mathbf{N}_{F(\rho)}^k(J)$. Once again the subscript denotes the ρ of the forest.

Training: In order to learn the structure of a tree we use the training set $\mathbf{I}$ and the distances $\rho(I_n, I_m)$ for each image pair in $\mathbf{I}$. Our goal is to find the sequence of binary tests on image features that sequentially partition $\mathbf{I}$ into the most spatially compact subsets with respect to $\rho(\cdot, \cdot)$. Assuming $\mathbf{I}$ is a representative dataset, the learned binary tests would then successfully apply to other images.

Given a node s and the set of training images at it, $\mathbf{I}_s$, we first define branching of s via the binary test and the partitioning of $\mathbf{I}_s$ into two as

$$t_s(I_n; m, \tau) \triangleq \begin{cases} I_n \in \mathbf{I}_{s_R}, & \text{if } f_T^m(I_n) > \tau, \\ I_n \in \mathbf{I}_{s_L}, & \text{if } f_T^m(I_n) \leq \tau, \end{cases} \quad \forall I_n \in \mathbf{I}_s \tag{1}$$

where f_T^m denotes the mth component of $\mathbf{f}_T(I_n)$, $\tau \in \mathbb{R}$, and s_L and s_R are the children of s. At every node we would like to optimize the parameters m and τ to obtain the most compact partitioning of $\mathbf{I_s}$. To do this we define spatial compactness of a set $\mathbf{A}$ with respect to $\rho(\cdot, \cdot)$ as

$$C_\rho(\mathbf{A}) \triangleq \frac{1}{|\mathbf{A}|^2} \sum_{I_i \in \mathbf{A}} \sum_{I_j \in \mathbf{A}} \rho(I_i, I_j), \tag{2}$$

where $|\mathbf{A}|$ denotes the size of the set and $C_\rho(\mathbf{A})$ its *cluster size*. Using $C_\rho(\cdot)$ we can formulate the gain in compactness a specific set of parameters yields with

$$G(\mathbf{I}_s, m, \tau) \triangleq C_\rho(\mathbf{I}_s) - \frac{|\mathbf{I}_{s_R}|}{|\mathbf{I}_s|} C_\rho(\mathbf{I}_{s_R}) - \frac{|\mathbf{I}_{s_L}|}{|\mathbf{I}_s|} C_\rho(\mathbf{I}_{s_L}), \tag{3}$$

where the weights $|\mathbf{I}_{s_R}|/|\mathbf{I}_s|$ and $|\mathbf{I}_{s_L}|/|\mathbf{I}_s|$ avoid constructing too small partitions. Using this formulation we determine the best possible binary test at node s with the following optimization problem

$$(m_s, \tau_s) = \arg_{m,\tau} \max \; G(\mathbf{I}_s, m, \tau). \tag{4}$$

In practice we do not take into account all m in the above optimization problem but choose a small random subset of the components of $\mathbf{f}_T(\cdot)$ at each node as is commonly done in decision forests [8]. The optimization over τ though is done through exhaustive search.

For each tree we start from its root node setting $\mathbf{I}_{s_0} = \mathbf{I}$. We then sequentially determine the binary tests using Eqn. 4 and add new nodes to the tree. We continue this process and grow the trees. The growth process is terminated at a node when i) we can no longer find a test that creates a more compact partitioning than the one in the node, i.e. $\forall(m, \tau), G < 0$, ii) the number of training images within the node is too small or iii) we reach at the maximum allowed depth and stop due to computational cost considerations.

3   Experiments

In this section we demonstrate NAF on two different applications. Our aim is to analyze NAF in different experimental setups and for different image spaces. We also highlight the application-specific components that can be changed to use NAF in different contexts. For both experiments we use 355 T1 weighted brain MR images from the publicly available OASIS dataset [12]. These images are skull stripped, histogram equalized and aligned to a common reference frame via affine registration. The resolution of each image is $1 \times 1 \times 1 \ mm^3$.

A. Choosing the Closest Images for Non-linear Registration: In the first application we focus on predicting the neighbourhood of a new image J within a dataset $\mathbf{I}$ with respect to the amount of deformation between images. We predict images in $\mathbf{I}$ that need the least amount of deformation to nonlinearly align them to J. This is a relevant problem for large cohort studies and multi-atlas based segmentation methods. Our aim in this experiment is to demonstrate the quality of NAF's predictions compared to the real neighbourhoods for this highly nonlinear problem. The application specific and experimental details are given below along with results and discussions.

$\rho(\cdot, \cdot)$: We measure the amount of deformation between two images using the distance $\rho(I, J) \triangleq \int_{\Omega_I} \log |\mathrm{Jac}(\Phi_{I \to J})| \, d\Omega_I + \int_{\Omega_J} \log |\mathrm{Jac}(\Phi_{J \to I})| \, d\Omega_J$, where Ω_I is the domain of I, $\Phi_{I \to J}$ is the deformation mapping I to J, i.e. $\Phi_{I \to J} \circ I = J$, and $\mathrm{Jac}(\cdot)$ is the Jacobian determinant. We use the diffeomorphic demons algorithm [18] for determining each deformation.

Dataset: The first 169 images are used in training and the rest 186 for testing.

Features: We randomly choose $Q = 10000$ pairs of voxels in the reference frame. Then we smooth each image with an averaging kernel of size $12 \times 12 \times 12 \ mm^3$. The feature vector for each image consists of the intensity differences between the pairs of voxels in the smooth version of the image.

NAF details: Using the training set we train a NAF of 1500 trees, each of maximum depth 6. Minimum number of allowable training images for a node is set to 7 beyond which we stop growing the tree. Each tree is constructed using a random subset of the entire feature vector of size $q = 1000$. For each test image J we predict its neighbourhood, $\mathbf{N}_{F(\rho)}^k(J)$, for different values of $k = 1, 3, 5, 7, 10$.

Evaluation and Results: For each test image J, we evaluate the quality of $\mathbf{N}_{F(\rho)}^k(J)$ by comparing it to the real neighbourhood $\mathbf{N}_\rho^k(J)$ using the following ratio

$$\varrho_J(\mathbf{N}_{F(\rho)}^k(J)) \triangleq \frac{\sum_{I \in \mathbf{N}_{F(\rho)}^k(J)} \rho(I, J)}{\sum_{I \in \mathbf{N}_\rho^k(\mathbf{J})} \rho(I, J)} \geq 1, \tag{5}$$

which measures how close the images in $\mathbf{N}_{F(\rho)}^k(J)$ to J compared to the ones in $\mathbf{N}_\rho^k(J)$. In Table 1 we provide the mean values and standard deviations of $\varrho_J(\mathbf{N}_{F(\rho)}^k(J))$ computed over 186 test images for different k. These values can be best interpreted in comparison with the ranges $\varrho_J(\cdot)$ can take for each k. In

order to present these ranges, for each test image J and each k we randomly chose 2000 subsets within the training set. We denote each of these subsets by $\mathbf{N}_{\mathbf{r}}^{k}(J)$. We then computed $\varrho_J(\mathbf{N}_{\mathbf{r}}^{k}(J))$ values and present the mean and standard deviations for these random subsets (computed over 186×2000 subsets for each k) in Table 1. Results given in Table 1 demonstrate that NAF predictions are indeed very close to the real neighbourhoods in terms of their distances to J. Especially in comparison with $\varrho_J(\mathbf{N}_{\mathbf{r}}^{k}(J))$ we notice that $\varrho_J(\mathbf{N}_{F(\rho)}^{k}(J))$ values are within the lowest part of the entire range of $\varrho_J(\cdot)$. We further plot in Figures 1(a) and (b) the normalized histograms for $\varrho_J(\mathbf{N}_{\mathbf{r}}^{k}(J))$ and $\mathbf{N}_{F(\rho)}^{k}(J)$ for $k = 1$ and $k = 7$. Comparing these histograms we see that the distribution of $\mathbf{N}_{F(\rho)}^{k}(J)$ is more concentrated close to one and it lies in the lower frequency region of the distribution for $\varrho_J(\mathbf{N}_{\mathbf{r}}^{k}(J))$. The difference is even more pronounced for $k = 7$, i.e. choosing multiple neighbours, which is more relevant for most applications such as multi-atlas based segmentation. Lastly in Figures 1(c)-(e) we show two sets of examples (different rows) where NAF predicts a different closest neighbour than the real one. However, visually the test image and the predicted neighbour are very similar.

Computation Times: For each test image NAF took at maximum 10.2 seconds to predict the neighbourhood with a C++ implementation on an Intel Xeon® at 2.27 GHz. Exhaustive search requires 169 nonlinear registrations which took on the average 1.9 hours for each test image.

B. Age Regression from Brain MR Scans: In the second application we focus on a high-dimensional image space equipped with a distance based on non-image based meta information: subject age. We devise an image-based regression algorithm powered by NAF to predict the age of a subject using the MR image. Our aim is to demonstrate the use of NAF for this type of applications and also quality of the predicted neighbourhood through an analysis end result.

$\rho(\cdot, \cdot)$: The distance of the image space is $\rho(I, J) = |\text{age}(I) - \text{age}(J)|$, where $\text{age}(I)$ denotes the subject's age with image I and $|\cdot|$ is the absolute value.

Dataset: We use the 355 images and perform leave-one-out tests.

Features: We randomly choose $Q = 10000$ voxels in the reference frame and use the intensity values taken from the images smoothed as in the previous case.

NAF details: Most details of NAF are the same as the previous case. The only differences is this time the maximum tree depth is 12 and we use 700 trees.

Evaluation and Results: In this application we evaluate NAF's results by comparing the real age of the test subject with the prediction obtained using the neighbourhood predicted by NAF. For each test image J we predict the age of the subject by taking the average age in $\mathbf{N}_{F(\rho)}^{15}(J)$. Figure 1(f) plots the predicted age vs. actual age for all 355 tests. The resulting correlation is reasonable high with a r-value $= 0.93$ ([17] reports slightly lower values for a slightly smaller dataset). We observe that NAF is able to approximate an informative image neighbourhood for a new image that is useful for the regression analysis.

Table 1. Top row: mean and standard deviations for the ratios of total distance from $\mathbf{N}_{F(\rho)}^k(J)$ to J and from $\mathbf{N}_\rho^k(J)$ to J, see Eqn. 5. Bottom row: presents the range of $\varrho_J(\cdot)$ within the training set by providing same values for random subsets of the training set. NAF predictions are very close the real neighbourhood considering the range of $\varrho_J(\cdot)$.

k	1	3	5	7	10
$\varrho_J(\mathbf{N}_{F(\rho)}^k(J))$	1.05 ± 0.04	1.05 ± 0.02	1.04 ± 0.02	1.04 ± 0.02	1.04 ± 0.01
$\varrho_J(\mathbf{N}_{\mathbf{r}}^k(J))$	1.20 ± 0.07	1.18 ± 0.06	1.18 ± 0.06	1.17 ± 0.06	1.16 ± 0.06
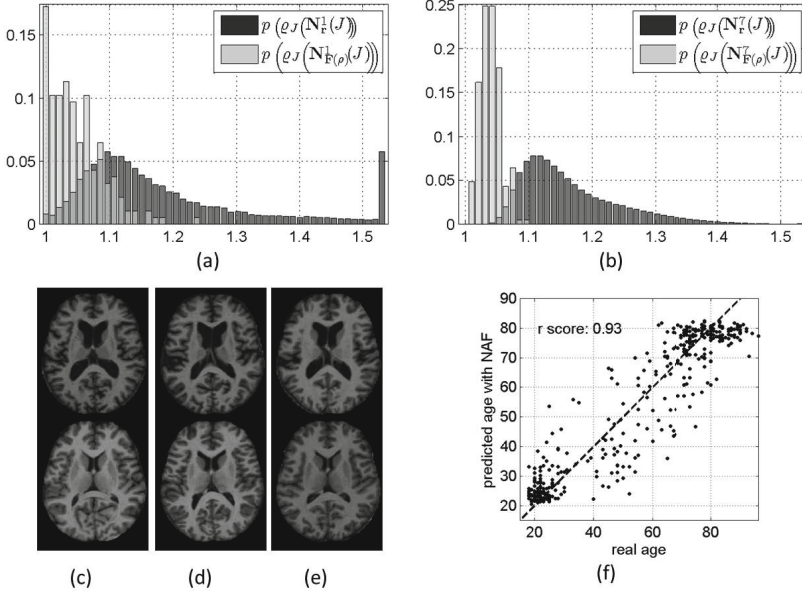
Fig. 1. Experiment A:(a,b) Normalized histograms of $\varrho_J(\mathbf{N}_{F(\rho)}^k(J))$ (light) and $\varrho_J(\mathbf{N}_{\mathbf{r}}^k(J))$ (dark) for $k = 1$, 7 respectively. NAF predictions are concentrated close to one and lie in the low frequency region of the distribution for $\varrho_J(\mathbf{N}_{\mathbf{r}}^k(J))$ (c)-(e) Two tests (different rows) where NAF suggests a different closest image than the real one: (c) the test image, (d) real closest (e) NAF prediction. Note that images are very similar visually. **Experiment B:**(f) Image-based regression for age prediction by NAF using $\mathbf{N}_{F(\rho)}^{15}(J)$. Note the high correlation $r = 0.93$.

4 Conclusion

We proposed an algorithm for solving one of the critical problems common to all neighborhood-based approaches for image analysis: approximating the neighborhood of a new image within a training set of images with respect to a given distance. The algorithm is general and can be applied to various tasks that utilize different distance definitions, as shown in the experiments. Furthermore, as the method is based on the framework of random decision forests the computation times are fast. We believe that applications such as multi-atlas registration and 'manifold'-based techniques can benefit from the proposed algorithm.

References

1. Aljabar, P., Heckemann, R., Hammers, A., Hajnal, J., Rueckert, D.: Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy. Neuroimage 46(3), 726–738 (2009)
2. Aljabar, P., Wolz, R., Rueckert, D.: Manifold Learning for Medical Image Registration, Segmentation, and Classification. In: Machine Learning in Computer-Aided Diagnosis: Medical Imaging Intelligence and Analysis. IGI Global (2012)
3. Allassonnire, S., Amit, Y., Trouv, A.: Towards a coherent statistical framework for dense deformable template estimation. J. R. Stat. Soc.: Series B 69 (2007)
4. Amit, Y., Geman, D.: Shape quantization and recognition with randomized trees. Neural Computation 9 (1997)
5. Bengio, Y., Paiement, J., Vincent, P., Delalleau, O., Le Roux, N., Ouimet, M.: Out-of-sample extensions for lle, isomap, mds, eigenmaps, and spectral clustering. In: NIPS, vol. 16 (2004)
6. Breiman, L.: Random forests. Machine Learning 45 (2001)
7. 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 (2011)
8. Criminisi, A., Shotton, J., Konukoglu, E.: Decision forests: A unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning. NOW Publishing: Foundations and Trends 7 (2012)
9. Gray, K.R., Aljabar, P., Heckemann, R.A., Hammers, A., Rueckert, D.: Random Forest-Based Manifold Learning for Classification of Imaging Data in Dementia. In: Suzuki, K., Wang, F., Shen, D., Yan, P. (eds.) MLMI 2011. LNCS, vol. 7009, pp. 159–166. Springer, Heidelberg (2011)
10. Hamm, J., Ye, D.H., Verma, R., Davatzikos, C.: GRAM: A framework for geodesic registration on anatomical manifolds. Med. Image Anal. 14 (2010)
11. Jia, H., Yap, P.T., Shen, D.: Iterative multi-atlas-based multi-image segmentation with tree-based registration. Neuroimage 59 (2012)
12. Marcus, D., Wang, T., Parker, J., Csernansky, J., Morris, J., Buckner, R.: Open access series of imaging studies (oasis): cross-sectional mri data in young, middle aged, nondemented, and demented older adults. J. of Cog. Neuroscience 19 (2007)
13. Nister, D., Stewenius, H.: Scalable recognition with a vocabulary tree. In: CVPR, vol. 2, pp. 2161–2168 (2006)
14. Niyogi, X.: Locality preserving projections. In: NIPS, vol. 16 (2004)
15. Norouzi, M., Fleet, D.: Minimal loss hashing for compact binary codes. In: ICML (2011)
16. Sabuncu, M.R., Balci, S.K., Shenton, M.E., Golland, P.: Image-driven population analysis through mixture modeling. IEEE Trans. Med. Imaging 28 (2009)
17. Sabuncu, M.R., Van Leemput, K.: The Relevance Voxel Machine (RVoxM): A Bayesian Method for Image-Based Prediction. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part III. LNCS, vol. 6893, pp. 99–106. Springer, Heidelberg (2011)
18. Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Diffeomorphic demons: Efficient non-parametric image registration. NeuroImage 45 (2009)
19. Weiss, Y., Torralba, A., Fergus, R.: Spectral hashing. In: NIPS (2008)
20. Wolz, R., Aljabar, P., Hajnal, J.V., Hammers, A., Rueckert, D., Weiner, M.W.: LEAP: learning embeddings for atlas propagation. Neuroimage 49 (2010)

Recognition in Ultrasound Videos: Where Am I?

Roland Kwitt[1], Nuno Vasconcelos[2], Sharif Razzaque[3], and Stephen Aylward[1]

[1] Kitware Inc., Carrboro, NC, USA
[2] Dept. of Electrical and Computer Engineering, UC San Diego, USA
[3] Dept. of Computer Science, UNC, Chapel Hill, USA

Abstract. A novel approach to the problem of locating and recognizing anatomical structures of interest in ultrasound (US) video is proposed. While addressing this challenge may be beneficial to US examinations in general, it is particularly useful in situations where portable US probes are used by less experienced personnel. The proposed solution is based on the hypothesis that, rather than their appearance in a single image, anatomical structures are most distinctively characterized by the variation of their appearance as the transducer moves. By drawing on recent advances in the non-linear modeling of video appearance and motion, using an extension of dynamic textures, successful location and recognition is demonstrated on two phantoms. We further analyze computational demands and preliminarily explore insensitivity to anatomic variations.

1 Motivation

Many developing countries, as well as rural areas of developed nations, do not have immediate access to expensive medical imaging equipment such as MR or CT. As highlighted in a recent study [8], ultrasound (US) imaging is particularly well-suited for those underserved areas, since it is low-cost, versatile and non-invasive. Additionally, medical care in emergency vehicles and military field operations may benefit from US when performing first care. However, as emphasized in [8], training is needed for personnel to unfold the full potential of US imaging; yet this training requirement can be problematic in rural and emergency situations due to cost and circumstance. In addition to US interpretation, high US acquisition quality is essential but often difficult to achieve.

A cost-effective and practical solution to the training challenge is to embed expertise into the US system and/or a connected mobile device. This has led other groups to attempt to embed computer-aided diagnosis (CAD) systems into imaging devices. We instead seek to help an inexperienced operator acquire *clinically significant* images that can then be transferred to a central location for reading by experts. This approach should be easier to achieve and more broadly useful than an embedded CAD system. The technical challenge reduces to developing algorithms for recognizing key anatomic structures as the US videos are acquired. Based on those localizations, the system can then convey to the operator how to acquire additional images, relative to those key locations, for transmission to expert readers, or can indicate when US video must be re-acquired to meet

(a) Data matrix   (b) Generative model of the KDT
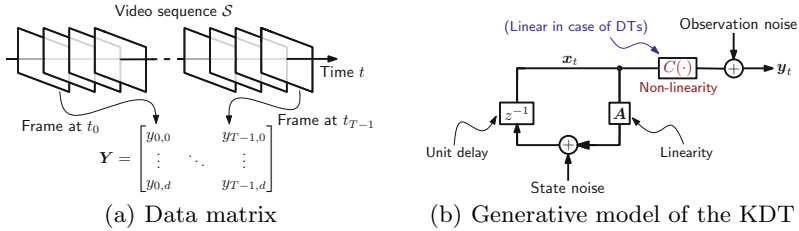
Fig. 1. (a) Assembly of the data matrix $\boldsymbol{Y}$ from a video sequence $\mathcal{S}$; (b) Generative model for the KDT

quality requirements. Apart from being helpful to novice users, we argue that automated recognition of anatomical structures might also be beneficial to experienced physicians, since it could help to minimize the risk of misidentifications.

The objective of this work is to promote a new pathway for locating anatomical structures when moving an US transducer. The key idea is to *avoid* image-to-image comparisons using an atlas but rather to exploit the full spatio-temporal information of the US video sequences. It is argued that the appearance changes of anatomical structures, due to probe motion, are particularly distinctive for their localization. Technically, we draw on recent advances in video modeling in computer vision. The varying appearance of an anatomical structure is represented by a generative video model, known as the *kernel dynamic texture* [2]. Similarity between video sequences is then defined as similarity in the parameter space of this model. Since we propose storing a database of key-location US sequences on portable devices and performing real-time analysis of US videos as they are acquired, generative models are particularly useful. In our case, we only need to augment the database by the KDT model parameters (which have a small memory footprint) and distances can be very efficiently computed.

While classification of US images has been previously studied (e.g., [7]), to the best of our knowledge, this is the first work to tackle localization on the basis of dynamic US sequence information. This paper presents 1) our application of the kernel dynamic texture algorithm, 2) a preliminary study on sensitivity and specificity using phantoms (admittedly for a limited range of the relevant problem space) and 3) a study on robustness towards simulated anatomic variations between the modeled structures to be localized and the actual observations.

2   Recognition with Kernel Dynamic Textures

We selected *dynamic texture (DT)* [5] models as an appropriate class of generative models for capturing video appearance changes. DT models arose from computer vision and were selected for US modeling because of the prominent role texture plays in US images, e.g., compared to edges or intensity. In particular, we exploited a recent non-linear extension of the DT family, denoted the *kernel dynamic texture (KDT)* [2], to capture non-linear appearance changes that will occur as structures move into and out of the ultrasound imaging plane.

Consider a US sequence $\mathcal{S}$ as an ordered sequence of T video frames, i.e., $\mathcal{S} = (\boldsymbol{y}_0, \ldots, \boldsymbol{y}_{T-1})$, where $\boldsymbol{y}_t \in \mathbb{R}^d$ is the frame observed at time t. Under the DT framework of [5], these observations are modeled as samples of a *linear dynamical system (LDS)*. At time t, a vector of state coefficients $\boldsymbol{x}_t \in \mathbb{R}^T$ is first sampled from a first-order Gauss-Markov process, and the state coefficients are then linearly combined into the observed video frame $\boldsymbol{y}_t$, according to

$$\boldsymbol{x}_t = \boldsymbol{A}\boldsymbol{x}_{t-1} + \boldsymbol{w}_t, \tag{1}$$

$$\boldsymbol{y}_t = \boldsymbol{C}\boldsymbol{x}_t + \boldsymbol{v}_t \tag{2}$$

where $\boldsymbol{A} \in \mathbb{R}^{T \times T}$ is the *state-transition* matrix and $\boldsymbol{C} \in \mathbb{R}^{d \times T}$ is the *generative* matrix that governs how the state determines the observation. Further, $\boldsymbol{w}_t$ and $\boldsymbol{v}_t$ denote state and observation noise with $\boldsymbol{w}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$ and $\boldsymbol{v}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R})$, respectively. Assuming that the observations are centered[1] and following the system identification strategy of [5], $\boldsymbol{C}$ is estimated by computing an SVD decomposition of the data matrix $\boldsymbol{Y} = [\boldsymbol{y}_0 \cdots \boldsymbol{y}_{T-1}]$ as $\boldsymbol{Y} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$ and setting $\boldsymbol{C} = \boldsymbol{U}$. The state matrix $\boldsymbol{X} = [\boldsymbol{x}_0 \cdots \boldsymbol{x}_{T-1}]$ is estimated as $\boldsymbol{X} = \boldsymbol{\Sigma}\boldsymbol{V}^\top$ and $\boldsymbol{A}$ can be computed using least-squares as $\boldsymbol{A} = [\boldsymbol{x}_1 \cdots \boldsymbol{x}_{T-1}][\boldsymbol{x}_0 \cdots \boldsymbol{x}_{T-2}]^\dagger$, where $\dagger$ denotes the pseudoinverse. When restricting the DT model to N states, $\boldsymbol{C}$ is restricted to the N eigenvectors corresponding to the N largest eigenvalues. The rest follows accordingly. Due to space limitations, refer to [5] for details on noise parameter estimation. In the non-linear DT extension of [2], the generative matrix $\boldsymbol{C}$ is replaced by a non-linear observation function $C : \mathbb{R}^T \to \mathbb{R}^d$, i.e.,

$$\boldsymbol{y}_t = C(\boldsymbol{x}_t) + \boldsymbol{v}_t, \tag{3}$$

while keeping the state evolvement linear. The corresponding dynamical system is denoted a *kernel dynamic texture (KDT)*, shown in Fig. 1(b). The non-linearity of C requires a different, although conceptually equivalent, set of parameter estimates. The idea is to use kernel PCA (KPCA) to learn the inverse mapping $D : \mathbb{R}^d \to \mathbb{R}^T$ from observation to state space, in which case the KPCA coefficients then represent the state variables.[2] We note that the KDTs are not necessarily restricted to work with intensity observation matrices; they will work with any kind of feature for which we can define a suitable kernel, c.f. [3].

Additionally, we have chosen to adopt the distance measure from [2] for measuring similarity of two video sequences, $\mathcal{S}^a$ and $\mathcal{S}^b$. This approach was chosen for its speed. It is based on an adaption of the Martin distance [6] among the corresponding DTs $\mathcal{D}_a = (\boldsymbol{A}_a, \boldsymbol{C}_a)$ and $\mathcal{D}_b = (\boldsymbol{A}_b, \boldsymbol{C}_b)$ with N states each. The (squared) Martin distance, given by [6,4]

$$d^2(\mathcal{S}^a, \mathcal{S}^b) = -\log \prod_{i=1}^N \cos^2(\phi_i), \tag{4}$$

is based on the subspace angles ϕ_i among the infinite observability matrices $\boldsymbol{O}_a$ and $\boldsymbol{O}_b$, defined as [4] $[\boldsymbol{C}_a^\top \ (\boldsymbol{C}_a\boldsymbol{A}_a)^\top \ (\boldsymbol{C}_a\boldsymbol{A}_a^2)^\top \cdots]^\top =: \boldsymbol{O}_a$. In fact, the $\cos(\phi_i)$ correspond to the N largest eigenvalues λ_i of the generalized eigenvalue problem

---

[1] Centering is straightforward by subtracting the column-wise means of $\boldsymbol{Y}$.
[2] See supp. material to [2] for centering in the feature space induced by the kernel.

Fig. 2. Illustration of the *noodle phantom*, made of gelatine and Soba noodles (left three images) and an abdominal CIRS phantom mounted in a water tank (right)

$$\begin{bmatrix} \mathbf{0} & \mathbf{O}_{ab} \\ \mathbf{O}_{ba} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} = \lambda \begin{bmatrix} \mathbf{O}_{aa} & \mathbf{0} \\ \mathbf{0} & \mathbf{O}_{bb} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \text{ with } \mathbf{O}_{ab} = \mathbf{O}_a^\top \mathbf{O}_b, \tag{5}$$

subject to $\mathbf{x}^\top \mathbf{O}_{aa} \mathbf{x} = 1$ and $\mathbf{y}^\top \mathbf{O}_{bb} \mathbf{y} = 1$. For DTs, computation of $\mathbf{O}_{ab}$ is straightforward, since the terms $\mathbf{C}_a^\top \mathbf{C}_b$ can be evaluated. For KDTs, it can be shown that computation of $\mathbf{C}_a^\top \mathbf{C}_b$ (which are no longer available) boils down to computing the inner products between the principle components of kernel matrix $K_{ij}^a = k(\mathbf{y}_i^a, \mathbf{y}_j^a)$ and $K_{ij}^b = k(\mathbf{y}_i^b, \mathbf{y}_j^b)$, i.e.,

$$\mathbf{O}_{ab} = \sum_{n=0}^{\infty} (\mathbf{A}_a^n)^\top \underbrace{\mathbf{C}_a^\top \mathbf{C}_b}_{\text{DTs}} \mathbf{A}_b^n \to \sum_{n=0}^{\infty} (\mathbf{A}_a^n)^\top \underbrace{\tilde{\boldsymbol{\alpha}}^\top \mathbf{G} \tilde{\boldsymbol{\beta}}}_{\text{KDTs}} \mathbf{A}_b^n, \tag{6}$$

where $\tilde{\boldsymbol{\alpha}} = [\tilde{\boldsymbol{\alpha}}_0 \cdots \tilde{\boldsymbol{\alpha}}_{T-1}]$, $\tilde{\boldsymbol{\beta}}$ are the (normalized) KPCA weight matrices with $\tilde{\boldsymbol{\alpha}}_i = \boldsymbol{\alpha}_i - 1/N(\mathbf{e}^\top \boldsymbol{\alpha}_i)\mathbf{e}$ and $\mathbf{G}$ is the kernel matrix with entries $G_{ij} = k(\mathbf{y}_i^a, \mathbf{y}_j^b)$. In the remainder of the paper, we use (4)-(6) for measuring similarity between US sequences and a standard RBF kernel for all kernel computations.[3]

For localization, we follow a sliding-window strategy, measuring how well a *key sequence* matches a subsequence from a long path (i.e., the *search sequence* $\mathcal{P}_n$) of acquisitions. That is, given Q frames in a key sequence, we move a sliding-window $\mathcal{W}_i$ of Q frames along a path by p frame increments. For each $\mathcal{W}_i$, we estimate the KDT parameters and compute the Martin distance to the KDT of the key sequence. A key sequence is indicated in a search sequence when the distance is minimal. At this time these minimums are illustrative. As more data and specific applications evolve, statistical likelihood methods will be used.

3   Experimental Protocol

For the studies in this paper, we use two different kinds of phantoms: 1) a homemade *noodle phantom* made of gelatine with embedded Soba noodles and 2) a triple modality 3D abdominal phantom (CIRS Model 057) mounted in a water tank, see Fig. 2. The noodle phantom is particularly useful, since the noodles are self-similar at a small scale, have ambiguous patterns of bends at

---

[3] For KPCA, kernel width is set as $\sigma^2 = \text{median}_{i,j}\|\mathbf{y}_i - \mathbf{y}_j\|^2$; to compute G_{ij}, it can be shown [2] that $\mathbf{y}_i^a$ and $\mathbf{y}_j^b$ need to be scaled by σ_a and σ_b first.

medium scales, and at large scales and in US sequences present a rich set of structures that are difficult to casually distinguish.

For imaging we use the Telemed LogicScan 128 INT-1Z kit. US frequency is set to 5Mhz. Penetration depth is 90mm on the noodle phantom and 150mm on the abdominal phantom. *Speckle reduction* is enabled in the US acquisition software. All images were acquired freehand, without tracking. We learn $N = 5$ state KDTs and clip each sequence to a central 300×300 (noodle phantom), or 200×200 (abdominal phantom) pixel window. Using more states did not lead to any improvements in the presented results.

3.1 Localization of Structures within US Sequences

The first experiment tested whether it is possible to localize key structures in the noodle phantom. Two different sets of acquisitions were made. The first set was composed of short (40 frames) US *key sequences* $\mathcal{S}_n$, captured by moving the US transducer over three different key structures to be localized, see Fig. 3. Key structures were chosen ad hoc by the probe operator. We then estimated KDT models for each of the three key structures using this first set of data. The second set was composed of longer US *search sequences* $\mathcal{P}_n$, acquired along multiple paths on the noodle phantom; these simulated searches for the key structures, see Fig. 4. On both sets, we tried to minimize probe tilt and rotation, but rotation and titling was inevitable. Note also that the acquisition direction of the key sequences matched the acquisition direction of the search sequences.

To evaluate *sensitivity*, we performed the localization using key sequences applied to search sequences that also covered the corresponding key structures. Distance plots are shown on the left-hand side of Fig. 5. To evaluate *specificity*, we repeated this experiment along multiple search paths that did *not* cover any of the key structures. Distance plots are on the right-hand side of Fig. 5.

To evaluate the robustness against shifts of the ultrasound imaging plane (e.g., partial inclusion of a key structure), we performed ten runs with random displacements δ_x, δ_y of the clipping window in x and y direction with $\delta_x, \delta_y \in \{-5, \ldots, 5\}$ pixel. Fig. 5 shows the Martin distance averaged over all clipping window positions for each sliding window index along three search paths (left). The enclosing light-blue hull illustrates the standard deviation.

Based on the above three experiments we make the following observations: 1) key structures exist at global minima in the Martin distance metric of a search sequence, when key structures are encountered; 2) Martin distance decreases as the sliding window moves towards a key structure and increases as it leaves the
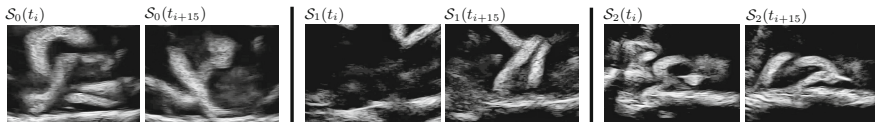
Fig. 3. Snapshots of three key structures at two time points on the noodle phantom

(a) Acquisition   (b) Overlap   (c) Movement
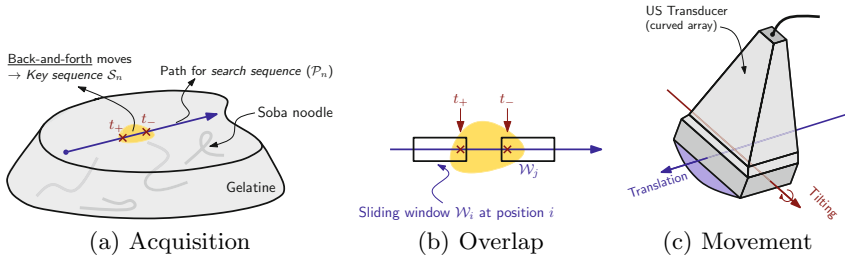
Fig. 4. Illustration of (a) the acquisition process on the noodle phantom, (b) sliding windows overlapping key structures (yellow) and (c) probe movements. In (b), hand-annotations t_+ and t_- bracket where the sliding window overlaps the key structure.
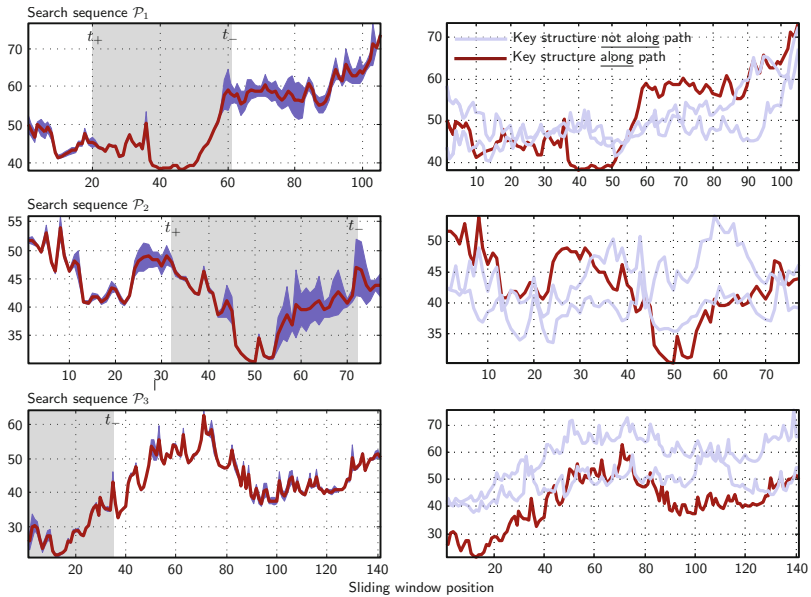
Fig. 5. Martin distance between key sequence KDT and the sliding window KTDs for three different paths, averaged over ten random clipping window positions (left); sliding window positions where the key structure is covered to some extent are marked light gray (from manual annotation); Distance measurements when trying to locate a key structure that was not covered by a path (right)

key structure; 3) if a key structure is not encountered by a search, then there is not a distinctive minimum in the distance measurements.

3.2   Localizing a Hepatic Vessel on an Abdominal Phantom

Our second experiment is more challenging in the sense that we try to locate a more subtle structure, namely a specific section of a *hepatic vessel* in an

abdominal phantom. The experimental protocol is similar to the previous experiment; however, US transducer *tilting* (also known as angulation [1]) (see Fig. 4) is used instead of translation along a path. We attempt to localize the hepatic vessel key structure within a single search sequence. The key sequence acquisition spans $\approx 40°$ around the angle where the vessel is visible. The search sequence covers $\approx 140°$ around the vessel. Again, all acquisitions were performed freehand, and the ultrasound probe was repositioned on the phantom between each acquisition. Fig. 6 shows the Martin distances for localization and for localization using shifted clipping windows.

This experiment highlights two things. First, we can again localize the key sequence within the longer search sequence, even though the span of the minimal Martin distances that correspond to the *true* location of the vessel is less prominent and less persistent than in the previous experiment. Second, variation in the distance measurements is much higher for small vessels than for the more distinct, larger structures form the noodle phantom.

3.3 Localization in the Presence of Simulated Anatomical Variation

Our third experiment focused on the insensitivity of the distance metric and the localization method to anatomic variations. Specifically, we simulate anatomic variation by (non-linear) spatial distortion of the search sequences. We admit
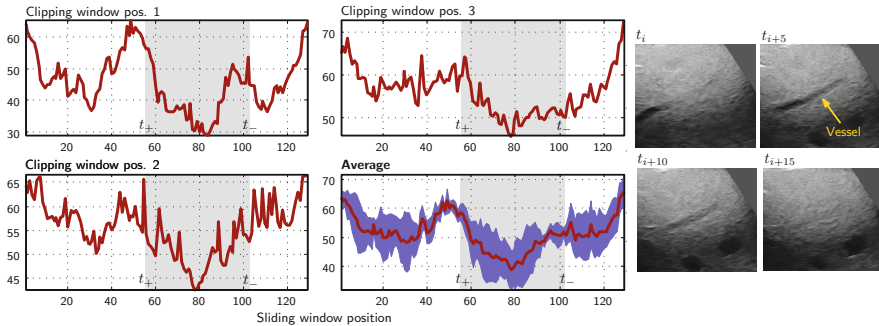
Fig. 6. Martin distance measurements for three clipping window positions and distance measurements averaged over all random runs (left); illustration of the hepatic vessel appearing and disappearing in the key sequence (right)
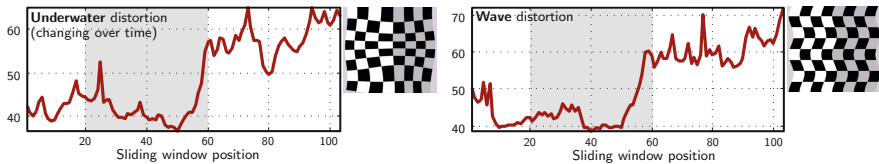
Fig. 7. Distortion experiments on the search sequence for two different types of (non-linear) spatial distortion (illustrated on the checkerboard pattern)

that this does not cover the range of variation among individuals, but it does begin to give an impression of robustness. Fig. 7 shows the distance measurements when the key structure is encountered within the search sequence. The distortions are illustrated on a checkerboard pattern. Note that the *underwater* distortion is changing over time.[4] As shown, the distortions do not have a negative impact on the localization, however, we observe less distinctive minima.

3.4 Discussion and Future Work

We conclude that the KDT framework + Martin distance is an effective combination in localizing short US sequences of key structures. Further, computational requirements are modest, even for our un-optimized (MATLAB) code and use of intensity features. For example, given a key sequence consisting of 40 frames, computing the KDT model and distance metric per sliding window requires ≈ 0.1 seconds on an Intel Core i7 1.7Ghz CPU with 8GB of RAM. It is also worth noting that we can perfectly recognize the short US key sequences using a simple nearest-neighbor classifier and the Martin distance as a metric. For future work we note that using intensity information as our observations space has its limitations. Due to the generic nature of the KDT approach and KPCA-based system identification, we could easily integrate more specifically tailored US features as long as we can define a suitable kernel.

References

1. Block, B.: The Practice of Ultrasound: A Step-by-Step Guide to Abdominal Scanning, 1st edn. Thieme (2004)
2. Chan, A., Vasconcelos, N.: Classifying video with kernel dynamic textures. In: CVPR, pp. 1–6 (2007)
3. Chaudry, R., Ravichandran, A., Hager, G., Vidal, R.: Histograms of oriented optical flow and binet-cauchy kernels on nonlinear dynamical systems for recognition of human actions. In: CVPR, pp. 1932–1939 (2009)
4. De Cock, K., Moore, B.: Subspace angles between linear stochastic models. In: CDC, pp. 1561–1566 (2000)
5. Doretto, G., Chiuso, A., Wu, Y., Soatto, S.: Dynamic textures. Int. J. Comput. Vision 51(2), 91–109 (2001)
6. Martin, R.: A metric for ARMA processes. IEEE Trans. Signal Process. 48(4), 1164–1170 (2000)
7. Sohail, A., Rahman, M., Bhattacharya, P., Krishnamurthy, S., Mudur, S.: Retrieval and classification of ultrasound images of ovarian cysts combining texture features and histogram moments. In: ISBI, pp. 288–291 (2010)
8. Spencer, J.: Utility of portable ultrasound in a community in Ghana. J. Ultrasound Med. 27(12), 1735–1743 (2008)

---

[4] See supp. video at http://vimeo.com/rkwitt

Self-similarity Weighted Mutual Information: A New Nonrigid Image Registration Metric

Hassan Rivaz and D. Louis Collins

Montreal Neurological Institute, McGill University

Abstract. Extending mutual information (MI), which has been widely used as a similarity measure for rigid registration of multi-modal images, to deformable registration is an active field of research. We propose a self-similarity weighted graph-based implementation of α-mutual information (α-MI) for nonrigid image registration. The new <u>Se</u>lf <u>Si</u>milarity $\underline{\alpha}$-<u>MI</u> (SeSaMI) metric takes local structures into account and is robust against signal non-stationarity and intensity distortions. We have used SeSaMI as the similarity measure in a regularized cost function with B-spline deformation field. Since the gradient of SeSaMI can be derived analytically, the cost function can be efficiently optimized using stochastic gradient descent. We show that SeSaMI produces a robust and smooth cost function and outperforms the state of the art statistical based similarity metrics in simulation and using data from image-guided neurosurgery.

1 Introduction

The joint intensity histogram of two images, of different or same modalities, is spread (i.e. the joint entropy is high) when they are not aligned, and is compact (i.e. the joint entropy is low) when the two images are aligned. Therefore, mutual information (MI) and its variants such as normalized MI (NMI) have been proposed and widely used for rigid registration of multi-modal images [1,2,3]. MI, being based on global intensity histograms, does not take into account local structures. Therefore, nonrigid registration, which has considerably more degrees of freedom and can distort local structures, is challenging with MI. It is also not robust against spatially varying bias fields. Exploiting the spatial information by conditioning MI calculation to location [4,5,6,7] has been shown to significantly improve nonrigid registration results.

In this work, we propose to incorporate image self-similarity into MI formulation. Self-similarity estimates the similarity of a point in one of the images to other points in the *same* image, and depends on local structures which are ignored by MI. Self-similarity was first proposed for object detection and image retrieval [8], and has since been used in image denoising [9] and registration [10].

Since self-similarity is calculated for pairs of points, it is natural to perceive it in a graph representation where image pixels are vertices and self-similarity is the weight of the edges. α-mutual information (α-MI) similarity
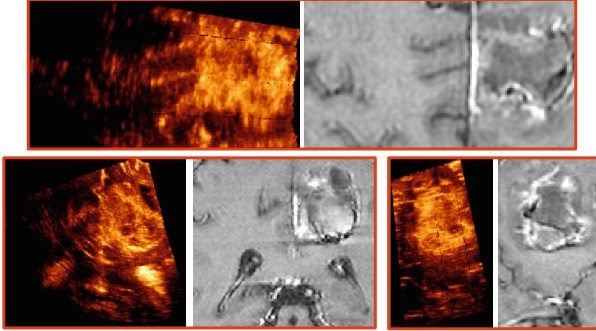
Fig. 1. Corresponding intra-operative US (hot colormap) and pre-operative MR (grayscale colormap) images of neurosurgery. A reconstructed US volume is sliced in the axial (top), coronal (bottom left) and sagittal (bottom right) directions. While local structures correspond, intensities are not related globally.

metric [11,12,13,14,15] is also graph based and has been recently shown to outperform MI in nonrigid registration applications. Therefore, we choose to incorporate self-similarity into this powerful registration framework. We apply the method to register pre-operative magnetic resonance (MR) images to intra-operative ultrasound (US) images in the context of image-guided neurosurgery (IGNS). Other works that nonlinearly register US to other modalities have used local correlation ratio [16] and MI of phase information [17]. Figure 1 shows an example of the registered US and MR images. The US images suffer from strong bias field due to signal attenuation, caused by scattering (from smaller than US wavelength inhomogeneities), specular reflection (from tissue boundaries) and absorption (as heat). In addition, US beam width varies significantly with depth, and therefore the same tissue looks different at different depths. Therefore, it is critical to exploit local structures.

In most image guided applications, one of the images is pre-operative. We therefore perform the self-similarity estimation only on this image offline, resulting in no increase in the on-line computational complexity. The pre-operative image is also usually of higher quality, making it a more attractive choice. We show that SeSaMI outperforms NMI and multi-feature α-MI in terms of producing a smooth dissimilarity function and registration accurary.

2   Rotationally Invariant Self-similarity Estimation

We first estimate a rotationally invariant 2D histogram descriptor for all pixels; such a sample pixel is marked with an asterisk in Figure 2 (we show 2D images for clarity; the arguments are trivially extended to 3D images). A circular patch with radius r is centered on the pixel, shown in the left image. In this work, we set r to 5 in 2D and to 2 in 3D. The axes of the histogram are d, the normalized Euclidian distance of the pixel from the center point, and i, the pixel's normalized intensity.
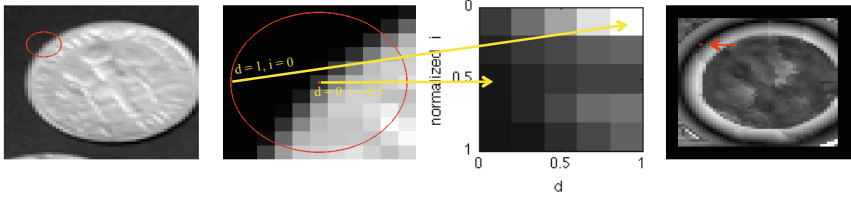
Fig. 2. Construction of the spin image and the resulting self-similarity distance. Left shows the image of a coin, with a circular patch of radius $r = 5$ pixels around a center point. Middle shows the zoomed patch, and the estimated 2D histogram descriptor (i.e. the spin image). Right shows the self-similarity distance to the center point (marked with an asterisk and pointed to by an arrow).

$d = 0$ and $d = 1$ in the histogram respectively correspond to the center pixel and to the pixels on the circle with radius $r = 5$ pixels. The intensity of pixels inside the patch is normalized to $[0\ 1]$. Each pixel inside the patch contributes to the 2D histogram: the histogram is constructed using a Gaussian Parzen window, i.e. a pixel with distance d to the center and normalized intensity i contributes to all bins according to a Gaussian centered at (d, i). The 2D histogram usually has a higher mass at higher d values because the number of pixels at distance d is proportional to d. Since d is the distance to the center (i.e. orientation is ignored), the 2D histogram descriptor is rotation invariant. It is also invariant to affine changes in the intensity because of the intensity normalization step. The histogram descriptor is similar to the spin image used in [18].

After calculating the 2D histogram descriptor for all points, we calculate the similarity between two points by the Earth Mover's Distance (EMD) [19]. The EMD metric avoids quantization and binning problems associated with histograms, and has been shown [19] to outperform other histogram comparison techniques. Figure 2 right shows the EMD distance to the point indicated by the asterisk. Note that small values of the EMD distance (darker pixels) represent more similar regions. It can be seen that the similarity metric is fully rotation invariant. We compute the self-similarity distance between each point and a window, of size The computational complexity of calculating the EMD distance is not an issue since it can be calculated offline on only the pre-operative image.

The histogram descriptor provides stability against small deformations of structures (due to the binning process), while subdividing the distance to the center (d in the histogram) encodes the spatial information. As a result, it is more robust than filter banks and differential invariants, which are also local descriptors [18]. Its disadvantage is its computational complexity. To reduce the running time and memory requirements, the EMD distance of a pixel with respect to pixels within its neighborhood is calculated. The EMD computation for a volume of size 100^3 pixels currently takes about 5 hours on a 3GHz processor.

The EMD distance provides a powerful metric to *condition* or *weight* MI estimation. *Conditioning* MI on EMD distance is motivated by the works which condition MI on spatial location and perform it on regions instead of the entire image [4,5,6,7]. Our preliminary results on conditioning MI on EMD have been

promising, especially since spatial distance can also be incorporated into the EMD distance. This is intuitive since self-similar pixels in one image are more likely to follow the same statistical relationship in the joint histogram. In this work, however, we focus on the second avenue, *weighting* MI with self-similarity. In the next section, we first briefly explain α-MI and then formulate SeSaMI.

3   Self-Similarity α-MI (SeSaMI)

Registration of a moving image I_m to a fixed image I_f can be formulated as

$$\hat{\boldsymbol{\mu}} = \arg\min_{\boldsymbol{\mu}} C, \quad C = S(I_f(\boldsymbol{x}), I_m(\mathbf{T}_{\boldsymbol{\mu}}(\boldsymbol{x})) + \frac{\omega_R}{2}\|\nabla\boldsymbol{\mu}\|^2 \tag{1}$$

where S is a dissimilarity metric, ω_R is a regularization penalty weight, ∇ is the gradient operator and $\mathbf{T}_{\boldsymbol{\mu}}$ is the transformation modeled by $\boldsymbol{\mu}$. We choose a free-form transformation parameterized by the location of cubic B-spline nodes. $\boldsymbol{\mu}$ is therefore a vector of the location of all the nodes in all directions.

MI similarity metric is usually calculated on the intensities only, and therefore the joint histogram is 2D. α-MI is usually calculated on multiple features like intensities and their gradients. Adopting the notation of [11], let $\boldsymbol{z}(\boldsymbol{x}_i) = [z_1(\boldsymbol{x}_i)\cdots z_q(\boldsymbol{x}_i)]$ be a q-dimensional vector containing all the features at point $\boldsymbol{x}_i$. Similar to [11], we choose image intensity and gradients at two different scales as features, resulting in 5 total features. Let $\boldsymbol{z}^f(\boldsymbol{x}_i)$ and $\boldsymbol{z}^m(\mathbf{T}_\mu(\boldsymbol{x}_i))$ be respectively the features of the fixed and moving image at $\boldsymbol{x}_i$ and $\mathbf{T}_\mu(\boldsymbol{x}_i)$, and $\boldsymbol{z}^{fm}(\boldsymbol{x}_i, \mathbf{T}_\mu(\boldsymbol{x}_i))$ be their concatenation $[\boldsymbol{z}^f(\boldsymbol{x}_i), \quad \boldsymbol{z}^m(\mathbf{T}_\mu(\boldsymbol{x}_i))]$. Minimal spanning tree (MST) [12] and k-nearest neighbor (kNN) [11,15] are among different methods for estimating α-MI from multi-feature samples. With N samples, the complexities of constructing MST and kNN graphs are $O(N^2 \log N)$ and $O(N \log N)$ respectively [14]. Therefore, we choose kNN.

Let $\boldsymbol{z}^f(\boldsymbol{x}_{ip})$, $\boldsymbol{z}^m(\mathbf{T}_\mu(\boldsymbol{x}_{ip}))$ and $\boldsymbol{z}^{fm}(\boldsymbol{x}_{ip}, \mathbf{T}_\mu(\boldsymbol{x}_{ip}))$ be respectively the nearest neighbors of $\boldsymbol{z}^f(\boldsymbol{x}_i)$, $\boldsymbol{z}^m(\mathbf{T}_\mu(\boldsymbol{x}_i))$ and $\boldsymbol{z}^{fm}(\boldsymbol{x}_i, \mathbf{T}_\mu(\boldsymbol{x}_i))$. Note that these three nearest neighbors in general do not correspond to the same point. To prevent notation clutter, we show the dependencies on location $\boldsymbol{x}_i$ or $\mathbf{T}_\mu(\boldsymbol{x}_i)$ only through i after this point whenever clear. Let $\boldsymbol{d}_{ip}^{fm} = \|\boldsymbol{z}_i^{fm} - \boldsymbol{z}_{ip}^{fm}\|$, and $\boldsymbol{d}_{ip}^m = \|\boldsymbol{z}_i^m - \boldsymbol{z}_{ip}^m\|$ ($\boldsymbol{d}_{ip}^{fm}$ and $\boldsymbol{d}_{ip}^m$ will be used later when we differentiate the cost function) and set

$$\Gamma_i^f = \sum_{p=1}^{k} \|\boldsymbol{z}_i^f - \boldsymbol{z}_{ip}^f\|, \ \ \Gamma_i^m(\boldsymbol{\mu}) = \sum_{p=1}^{k} \|\boldsymbol{d}_{ip}^m\|, \ \ \Gamma_i^{fm}(\boldsymbol{\mu}) = \sum_{p=1}^{k} \|\boldsymbol{d}_{ip}^{fm}\|. \tag{2}$$

A kNN estimator for α-MI$=-S$ (the dissimilarity function in Eq. 1) is

$$\widehat{\alpha\text{-MI}}(\boldsymbol{\mu}) = \frac{1}{\alpha - 1} \log \frac{1}{N^\alpha} \sum_{i=1}^{N} \left(\frac{\Gamma_i^{fm}(\boldsymbol{\mu})}{\sqrt{\Gamma_i^f \Gamma_i^m(\boldsymbol{\mu})}} \right)^{2\gamma} \tag{3}$$

where $\gamma = (1 - \alpha)q$ and $0 < \alpha < 1$; experimental results of rigid registration in [12] suggest that for MST graphs, α close to 1 gives better registration accuracy, while α close to 0.5 yields a wider capture range. We set $\alpha = 0.99$ in this work.

Weighting α-MI by Self-similarity. In an analogy to MI, small Γ_i^{fm} for majority of locations i means that data in the joint histogram is clustered and compact, and Γ_i^f and Γ_i^m are for normalization. Therefore, accurate estimates of Γ_i^{fm} are essential. Generally, most of the nearest neighbors in the joint feature space are also the most self-similar. However, due to spatially varying bias, small geometrical distortions, lack of enough number of features and misalignment, not all the nearest neighbors are self-similar. Therefore, to penalize points that are close but are not self-similar, we modify Γ_i^{fm} by

$$\Gamma_i^{fm}(\boldsymbol{\mu}) = \sum_{p=1}^{k} w_{ip} \|\boldsymbol{d}_{ip}^{fm}\|, w_{ip} = \text{EMD}(H(\boldsymbol{x}_i), H(\boldsymbol{x}_{ip})) \tag{4}$$

where $\text{EMD}(H(\boldsymbol{x}_i), H(\boldsymbol{x}_{ip}))$ is the EMD between the histogram descriptors.

We adopt an iterative stochastic gradient descent optimization method [20] for solving Eq. 1, which is fast and is less likely to get trapped in local minima. Therefore, $\boldsymbol{\mu}_{t+1} = \boldsymbol{\mu}_t + a_t \nabla_{\boldsymbol{\mu}} C$ where $\nabla_{\boldsymbol{\mu}} C$ is the gradient of C (from Eq. 1) wrt $\boldsymbol{\mu}$. The step size is a decaying function of the iteration number: $a_t = a/(A+t)^{\tau}$, with $a > 0, A \geq 0$ and $0 < \tau \leq 1$ user-defined constants [20]. From Eq. 1, we have $\nabla_{\boldsymbol{\mu}} C = -\nabla_{\boldsymbol{\mu}} \widehat{\alpha\text{-MI}} + \omega_R \Delta \boldsymbol{\mu}$ where $\Delta = \nabla.\nabla$ is the Laplacian operator. At a specific $\boldsymbol{\mu}$ where the graph topology changes, Eq. 3 can be non-differentiable [12]. However, assuming the topology does not change for small changes in $\boldsymbol{\mu}$, the gradient of $\widehat{\alpha\text{-MI}}$ is calculated analytically in [11,15] using the chain rule; due to space limitations, we refer the reader to them for details. The chain rule finally results in computation of the $\nabla_{\boldsymbol{\mu}} \Gamma_i^{fm}(\boldsymbol{\mu})$. From Eq. 4, we have

$$\frac{\partial}{\partial \mu_j} \Gamma_i^{fm}(\boldsymbol{\mu}) = \sum_{p=1}^{k} \frac{w_{ip}}{\|\boldsymbol{d}_{ip}^{fm}\|} \boldsymbol{d}_{ip}^{fmT} \cdot \frac{\partial}{\partial \mu_j} \boldsymbol{d}_{ip}^{fm} = \sum_{p=1}^{k} \frac{w_{ip}}{\|\boldsymbol{d}_{ip}^{fm}\|} \boldsymbol{d}_{ip}^{mT} \cdot \frac{\partial}{\partial \mu_j} \boldsymbol{d}_{ip}^{m} \tag{5}$$

where T means transpose. w_{ip} is calculated for either I_f or I_m; in the former case, its derivative wrt $\boldsymbol{\mu}$ is trivially zero, and in the latter case it is zero because the 2D histogram descriptors are invariant to small deformations [18]. Also, even for large global deformations, the histogram patches can be assumed locally rigid. The second equality is true because $\boldsymbol{z}^{fm}$ is the concatenation of $\boldsymbol{z}^f$ and $\boldsymbol{z}^m$, and $\partial \boldsymbol{z}^f / \partial \boldsymbol{\mu} = 0$. Finally,

$$\boldsymbol{d}_{ip}^{mT} \cdot \frac{\partial}{\partial \mu_j} \boldsymbol{d}_{ip}^{m} = \boldsymbol{d}_{ip}^{mT} \cdot \left(\frac{\partial}{\partial \mathbf{T}(\boldsymbol{x}_i)} \boldsymbol{z}_i^m \frac{\partial}{\partial \mu_j} \mathbf{T}(\boldsymbol{x}_i) - \frac{\partial}{\partial \mathbf{T}(\boldsymbol{x}_{ip})} \boldsymbol{z}_{ip}^m \frac{\partial}{\partial \mu_j} \mathbf{T}(\boldsymbol{x}_{ip}) \right). \tag{6}$$

Note that partial derivative of $\boldsymbol{z}^m$ wrt $\mathbf{T}$ means calculating derivatives in I_m's native space, i.e. wrt its own x, y or z coordinate. In our implementation, we precompute all the features of the I_f and I_m, and the derivatives of I_m's features wrt x, y and z directions.

(a) Biased red (I_f)   (c) α-MI-5f   (e) SeSaMI-5f

(g) NMI

(b) Green (I_m)   (d) α-MI-3f   (f) SeSaMI-3f
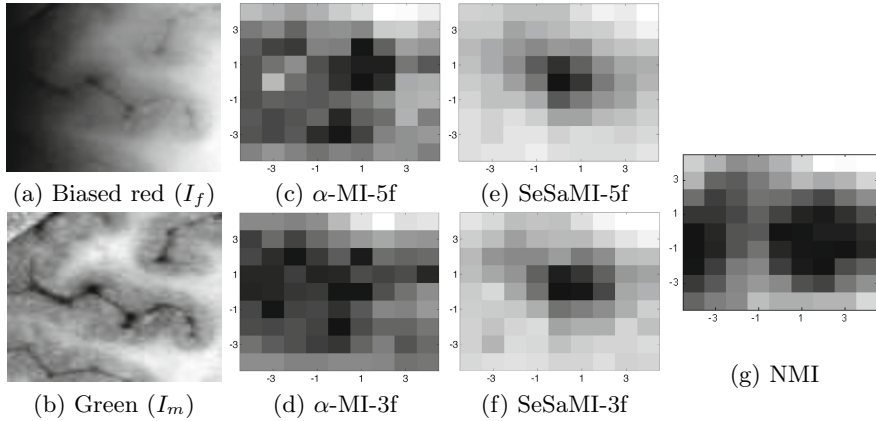
Fig. 3. Effect of the bias on the dissimilarity metrics in the human brain images. (a) The red channel with an additive bias. (b) The green channel. (c)-(f) The αMI and SeSaMI dissimilarity metrics calculated from $N = 400$ points randomly selected throughout I_f with 5 or 3 features (5f or 3f). (g) Our NMI implementation. The x and y axis represent the amount of rigid displacements of I_m in those directions (maximum of ± 4 pixels). Images are registered at 0 displacement. The self-similarity metrics in (e)-(f) are calculated using the biased I_f.

4   Experiments and Results

Visible Human Project. We test the new similarity metric on red and green channels of the visible human project, which are intrinsically registered. The data is publicly available at www.nlm.nih.gov/research/visible/ visible_human.html. We set the red image as I_f and add bias to it to show the robustness of our self-similarity measure and SeSaMI (the self-similarity is calculated on the biased I_f image). Figure 3 shows the results; in (c)-(f) a total of $N = 400$ randomly selected points are used (the same random points are used in all the 4 cases) and the number of nearest neighbors k is 100. (g) is our NMI implementation with Parzen window histogram estimation [3]. SeSaMI successfully gives the global optimum at 0, and also produces a relatively smooth dissimilarity metric. In addition, it gives a global minimum even with 3 features (original intensity, smoothed intensity and gradient magnitude), instead of 5 features (calculated at 2 scales). Reduction of the number of features makes the algorithm run faster.

US and MR. We apply our registration algorithm to the clinical data of the IGNS obtained from 13 patients with gliomas in the Montreal Neurological Institute. The pre-operative MR images are gadolunium-enhanced T1 weighted and are acquired approximately 2 weeks before the surgery. The intra-operative US images are obtained using an HDI 5000 (Philips, Bothell, WA) with a P7-4 MHz

Table 1. MR/US registration mTRE (mm) for 3 nonlinear registration methods

Patient	No. of landmarks	Initial	NMI	α-MI	SeSaMI
P1	35	6.30	11.93	2.32	2.05
P2	40	9.38	19.36	3.14	2.76
P3	32	3.93	13.43	1.83	1.92
P4	31	2.62	18.82	2.62	2.71
P5	37	2.30	15.76	1.97	1.89
P6	19	3.04	9.01	2.28	2.05
P7	23	3.75	16.03	3.05	2.89
P8	21	5.09	7.83	2.44	2.93
P9	25	3.00	14.05	2.83	2.75
P10	25	1.52	18.65	1.44	1.28
P11	21	3.70	11.01	2.81	2.67
P12	23	5.15	17.46	3.37	2.82
P13	23	3.78	9.15	2.45	2.34

phased array transducer. The ultrasound probe is tracked with a Polaris camera (NDI, Waterloo, Canada), and 3D US volumes are reconstructed using the tracking information. The tracking information is also used to perform the initial rigid registration of MR to US; a sample of this initial registration is shown in Figure 1. A neurosurgeon and two experts have selected corresponding anatomical landmarks in US and MR images in sub-voxel accuracy, which are used to calculate mTRE for validation. Table 1 shows that multi-feature α-MI and SeSaMI significantly outperform NMI in nonlinear registration of MR to US in all the 13 cases. In 10 out of 13 cases, SeSaMI gives the most accurate results due to its robust self-similarity measure incorporated into the powerful multi-feature α-MI similarity metric.

5   Conclusions

We introduced SeSaMI, a similarity metric that incorporates rotation and bias invariant self-similarity measures into graph-based α-MI. SeSaMI exploits self-similarity in a kNN α-MI registration framework by penalizing clusters (i.e. the nearest neighbors) that are not self-similar. Therefore, it significantly reduces the number of incorrect local minima as shown in Figure 3. We have also, for the first time, shown that multi-feature α-MI and SeSaMI significantly increase the registration accuracy of MR to US registration in our on-going IGNS project. In the future we will investigate GPU implementations of SeSaMI to achieve our goal of near-real time intra-operative US-MRI registration.

References

1. Wells, W., Viola, P., Atsumid, H., Nakajimae, S., Kikinis, R.: Multi-modal volume registration maximization of mutual information. Med. Imag. Anal. 1, 35–51 (1996)
2. Maes, F., et al.: Multimodality image registration by maximization of mutual information. IEEE Trans. Medical Imag. 16, 187–198 (1997)
3. Pluim, J., Maintz, J., Viergever, M.: Mutual-information-based registration of medical images: a survey. IEEE Trans. Medical Imag. 22, 986–1004 (2003)
4. Studholme, C., Drapaca, C., Iordanova, B., Cardenas, V.: Deformation-based mapping of volume change from serial brain MRI in the presence of local tissue contrast change. IEEE Trans. Medical Imag. 25, 626–639 (2006)
5. Loeckx, D., Slagmolen, P., Maes, F., Vandermeulen, D., Suetens, P.: Nonrigid image registration using conditional mutual information. IEEE Trans. Medical Imag. 29, 19–29 (2010)
6. Klein, S., et al.: Automatic segmentation of the prostate in 3D MR images by atlas matching using localized mutual information. Med. Phys. 35, 1407–1417 (2008)
7. Zhuang, S., Arridge, D., Hawkes, D., Ourselin, S.: A nonrigid registration framework using spatially encoded mutual information and free-form deformations. IEEE Trans. Medical Imag. (in press)
8. Shechtman, E., Irani, M.: Matching local self-similarities across images and videos. In: Computer Vision and Pattern Recognition (CVPR), pp. 1–8 (2007)
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. Imag. 27, 425–441 (2008)
10. Heinrich, M.P., Jenkinson, M., Bhushan, M., Matin, T., Gleeson, F.V., Brady, J.M., Schnabel, J.A.: Non-local Shape Descriptor: A New Similarity Metric for Deformable Multi-modal Registration. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part II. LNCS, vol. 6892, pp. 541–548. Springer, Heidelberg (2011)
11. Staring, M., Heide, U., Klein, S., Pluim, J.: Registration of cervical MRI using multifeature mutual information. IEEE Trans. Medical Imag. 28, 1412–1421 (2009)
12. Sabuncu, M., Ramadge, P.: Using spanning graphs for efficient image registration. Information Processing Medical Imag. (IPMI) 17, 788–797 (2008)
13. Kybic, J., Vnucko, I.: Approximate all nearest neighbor search for high dimensional entropy estimation for image registration. Signal Processing 92, 1302–1316 (2012)
14. Neemuchwala, H., Hero, A.: Entropic graphs for registration. In: Blum, R., Liu, Z. (eds.) Multi-sensor Image Fusion and its Applications. CRC Press (2005)
15. Oubel, E., Craene, M., Hero, A., Frangi, A.: Cardiac motion estimation by joint alignment of tagged MRI sequences. Med. Imag. Anal. 16, 339–350 (2012)
16. Wein, W., et al.: Automatic ct-ultrasound registration for diagnostic imaging and image-guided intervention. Medical Imag. Analysis 12, 577–585 (2008)
17. Zhang, W., Brady, M., Becher, H., Noble, A.: Spatio-temporal (2d+t) non-rigid registration of real-time 3D echocardiography and cardiovascular MR image sequences. Physics Med. Biol. 56, 1341–1360 (2011)
18. Lazebnik, S., Schmid, C., Ponce, J.: A sparse texture representation using local affine regions. IEEE Trans. Pattern Anal. Machine Int. 27, 1265–1278 (2005)
19. Rubner, Y., Tomasi, C., Guibas, L.: The earth mover's distance as a metric for image retrieval. IEEE Trans. Pattern Anal. Machine Int. 40, 99–121 (2000)
20. Klein, S., Staring, M., Pluim, J.: Evaluation of optimization methods for nonrigid medical image registration using mutual information and b-splines. IEEE Trans. Imag. Proc. 16, 2879–2890 (2007)

Inter-Point Procrustes: Identifying Regional and Large Differences in 3D Anatomical Shapes

Karim Lekadir[1], Alejandro F. Frangi[1], and Guang-Zhong Yang[2]

[1] Center for Computational Imaging & Simulation Technologies in Biomedicine
Universitat Pompeu Fabra and CIBER-BBN, Barcelona, Spain
[2] Hamlyn Centre for Robotic Surgery, Imperial College London, United Kingdom

Abstract. This paper presents a new approach for the robust alignment and interpretation of 3D anatomical structures with large and localized shape differences. In such situations, existing techniques based on the well-known Procrustes analysis can be significantly affected due to the introduced non-Gaussian distribution of the residuals. In the proposed technique, influential points that induce large dissimilarities are identified and displaced with the aim to obtain an intermediate template with an improved distribution of the residuals. The key element of the algorithm is the use of pose invariant shape variables to robustly guide both the influential point detection and displacement steps. The intermediate template is then used as the basis for the estimation of the final pose parameters between the source and destination shapes, enabling to effectively highlight the regional differences of interest. The validation using synthetic and real datasets of different morphologies demonstrates robustness up-to 50% regional differences and potential for shape classification.

1 Introduction

Procrustes analysis is a well-established approach for the alignment and interpretation of sets of corresponding landmark shapes [1]. Statistical shape modeling [2], shape classification [3], and regional shape analysis [4] are amongst the most common applications. In medical imaging, the shape complexity and the high variability means the 3D alignment needs to be robust and consistent in order to allow for a biologically meaningful analysis. It is now accepted that the standard Procrustes with its least square minimization can introduce errors in the presence of large shape differences [3, 5]. This is particularly the case when the shape variation is localized in a region of the anatomy, thus generating a non-Gaussian distribution of the residuals in both the spatial and frequency domains. Such situation is common in medical imaging, for instance due to abnormal remodeling at specific regions of the anatomy [4]. Thus far, the most common strategy to improve the Procrustes analysis is by using robust statistics (such as by using M-estimators [6]) but the performance is limited for regional shape analysis. More recently, two approaches have been proposed to increase the robustness of the Procrustes alignment. Firstly, the method in [7] replaces the least squares criterion with the more robust L_1-norm, but the technique is only valid in 2D, while anatomical shapes are mostly studied in 3D and increasingly in 4D.

A supervised Procrustes was also presented in [3] with the aim to aid classification, but the method requires a potentially biased or unavailable prior input.

In this paper, we propose a new Procrustes method that is unsupervised, robust and can be applied to different morphology types. To this end, instead of introducing a new similarity or minimization scheme, the proposed technique uses pose invariant shape variables to robustly compare the landmark configurations and to generate an intermediate template that is used for a geometrically meaningful alignment. More specifically, the influential landmarks (*i.e.*, those corresponding to large residuals) are identified using an invariant measure of dissimilarity and thus independently of the initial pose. Subsequently, these points are displaced within the source shape such that the new invariant discrepancies are minimized. The intermediate template as compared to the destination shape is free of large landmark residuals, thus enabling a successful final Procrustes alignment of the two input shapes. Detailed synthetic and real data experiments with various morphologies are carried out to assess the performance of the introduced algorithm and its value for shape discrimination.

2   Methods

2.1   Invariant Detection of Influential Points

The presence of regional anatomical differences means a subset of the landmarks carry a large amount of the total shape variability. Conventionally, the identification and handling of these influential points is achieved by performing an initial Procrustes alignment, followed by an iterative weighting of the landmarks based on the obtained landmark residuals. However, when the differences are large and localized, this initial least squares Procrustes becomes significantly and irreversibly corrupted, which does not allow for accurate weighting of the landmarks at subsequent iterations. To tackle this problem, we present in this work a new procedure for influential landmark identification that is independent of any pose estimation. To this end, variables that exclusively describe shape information are used. Let us denote $\mathbf{x}^{(s)}$ and $\mathbf{x}^{(d)}$ the source and destination shape vectors, respectively, encapsulating the coordinates of n corresponding points ($\mathbf{x}_i = (x_i, y_i, z_i)^T$ ($1 \leq i \leq n$)). For the proposed technique, each shape is associated with an inter-landmark distance matrix $\mathbf{D}$ of size $n \times n$, *i.e.*, :

$$\mathbf{D} = \begin{bmatrix} 0 & d_{12} & \dots & \dots & d_{1n} \\ d_{21} & 0 & \dots & \dots & d_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ d_{n1} & d_{n2} & \dots & \dots & 0 \end{bmatrix} \tag{1}$$

where $d_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|_2$ is the Euclidean distance between points p_i and p_j. Such shape variables have the advantage over point coordinates to be invariant to both

translation and rotation of the shapes. Furthermore, they can be used to robustly correct for size differences, by applying to $\mathbf{x}^{(s)}$ the following robust scaling factor:

$$\underset{1\le i<j\le n}{\mathrm{med}}\left(\frac{d_{ij}^{(d)}}{d_{ij}^{(s)}}\right). \tag{2}$$

The aim of the proposed method is to derive an invariant and robust measure of shape dissimilarity at each landmark point based on the differences in the inter-landmark distances calculated from the two shapes. Typically, a landmark associated with a significant shape difference will have many large differences of corresponding inter-landmark distances. Similarly, points corresponding to smaller shape residuals will have a large number of small inter-landmark differences. As a result, a median based measure of shape dissimilarity at each landmark is introduced as follows:

$$\delta_i = \underset{1\le j\le n}{\mathrm{med}}\left[(d_{ij}^{(s)} - d_{ij}^{(d)})^2\right]. \tag{3}$$

δ_i gives a good indication of the type of histogram involved with the inter-landmark differences for the point p_i. Subsequently, the set A of influential points is expected to be associated with particularly high δ_i values. Based on robust statistics [8], an appropriate definition of A can be derived by using robust estimations of the average point dissimilarity and deviation, $i.e,$:

$$A = \left\{ p_k \mid \delta_k > c(1.4826 \underset{1\le i\le n}{\mathrm{med}} \left| \delta_i - \delta^* \right|) \right\} \text{ where } \delta^* = \underset{1\le i\le n}{\mathrm{med}}(\delta_i). \tag{4}$$

The parameter c describes the number of deviations from the central value to consider and is typically chosen between 2 and 3. Next section describes how the identified shape dissimilarities are taken into account to derive a more robust and geometrically meaningful alignment.

2.2   Iterative Displacement of Influential Points

The aim of the proposed algorithm is to derive an intermediate template with an improved distribution of landmarks residuals. It will be subsequently used as a link between the source and destination shapes to enable robust alignment. The intermediate template vector is denoted as $\hat{\mathbf{x}}$ (its inter-landmark-distances as $\hat{d}_{ij}$) and it is initialized with the source shape vector $\mathbf{x}^{(s)}$. To obtain $\hat{\mathbf{x}}$, each influential point $p_k \in A$ is displaced with the aim to minimize the associated large residual by approaching the invariant properties of $\mathbf{x}^{(d)}$. To this end, a displacement vector $d\hat{\mathbf{x}}_k$ is calculated such that it minimizes the sum of the associated inter-landmark differences, $i.e.,$

$$d\hat{\mathbf{x}}_k = \underset{d\hat{\mathbf{x}}_k}{\arg\min} \sum_{j\in B} (\hat{d}_{kj}(\hat{\mathbf{x}}_k + d\hat{\mathbf{x}}_k) - d_{kj}^{(d)})^2 \tag{5}$$
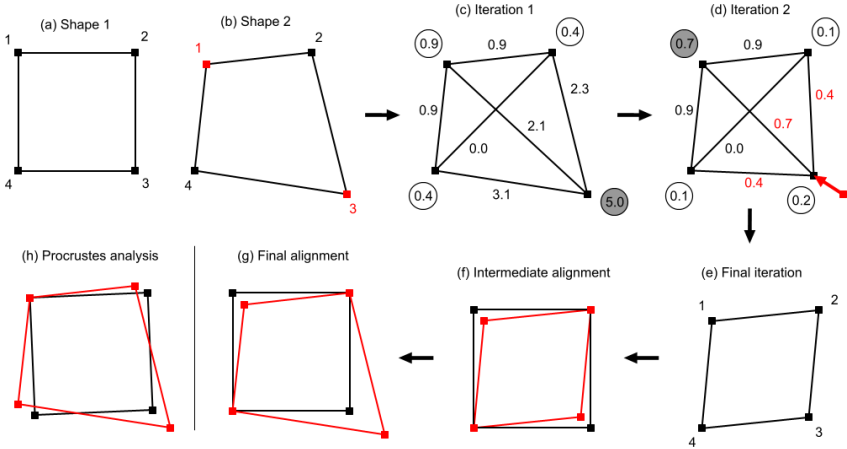
Fig. 1. Toy problem illustrating the main stages involved in the proposed technique. (a) and (b) show two quadrilaterals with differences at landmark 1 and 3. The calculated inter-landmark differences (see edges) and the associated landmark dissimilarity measures (Eq. (3), see circled) are displayed in (c) for iteration 1. The landmark 3 is identified as an influential point and displaced as shown in (d). It can be noticed how the associated inter-landmark discrepancies are decreased. At iteration 2 (d), the landmark 1 becomes an influential point and the algorithm continues until the intermediate template is obtained in (e). It can be seen how the transformation from (c) to (e) allows a balanced intermediate Procrustes alignment with the destination shape (f). The final result in (g) demonstrates good identification of the expected residuals. The Procrustes analysis alone in (h), however, introduces errors at all landmarks.

where B denotes the set of non-influential landmarks in the shape (*i.e.*, all landmarks in the shape minus A). By differentiation of (5), it can be easily shown that an improved $d\hat{\mathbf{x}}_k$ can be calculated at time $t+1$ based on the following formulae:

$$d\hat{\mathbf{x}}_k(t+1) = \alpha_k \sum_{j \in S} (\hat{d}_{kj}(t) - d_{kj}^{(d)}) \frac{(\hat{\mathbf{x}}_k(t) - \hat{\mathbf{x}}_j)}{\hat{d}_{kj}(t)}. \tag{6}$$

The parameter α_k is the optimal displacement step which can be found through simple line minimization. It is worth noting that $d\hat{\mathbf{x}}_k$ corresponds to a weighted sum of the unit vectors between $\hat{\mathbf{x}}_k$ and each non-influential point in B, where the weights $\hat{d}_{kj} - d_{kj}^{(d)}$ promote displacements in the directions corresponding to large inter-landmark discrepancies. The vector $d\hat{\mathbf{x}}_k$ is typically obtained after a number of successive displacements following Eq. (6). The inter-landmark distance matrices $\hat{\mathbf{D}}$ and $\mathbf{D}^{(d)}$ are subsequently updated, followed by a new round of influential point detection and displacement. This iterative approach is a key element of the proposed technique as it allows identifying both large and less significant shape differences. The most severe landmark residuals are typically detected during the initial iterations. After their correction, their effects are eliminated which facilitates the identification of less severe influential landmarks at subsequent iterations. The iterative procedure

then is repeated until no influential point remains, indicating that the shape differences between $\hat{\mathbf{x}}$ and $\mathbf{x}^{(d)}$ become normally distributed along all landmarks. Finally, a Procrustes alignment is carried out between the intermediate and destination shapes, followed by the application of the corresponding parameters to the original vector $\mathbf{x}^{(s)}$ for final alignment with $\mathbf{x}^{(d)}$.

It is worth mentioning that unlike existing techniques, the inter-point Procrustes has an interpretative power since it explicitly distinguishes between common structures and dissimilar regions in the shape. The proposed algorithm is now summarized in Table 1 and illustrated in the synthetic example of Fig. 1.

Table 1. Listing for the proposed alignment algorithm

1.	$\mathbf{x}^{(s)}$ and $\mathbf{x}^{(d)}$ are the input shape vectors. $\hat{\mathbf{x}}$ is the output intermediate vector initialized as $\hat{\mathbf{x}} = \mathbf{x}^{(s)}$.
2.	Calculate the inter-landmark distance matrices $\hat{\mathbf{D}}$ and $\mathbf{D}^{(d)}$.
3.	Calculate the median discrepancy values δ_i.
4.	Identify the set of influential landmarks A (Eq. (4)).
5.	If $A = \varnothing$ go to step (**8**).
6.	Displace the influential points in A (Eq. (6)).
7.	Go to step (**2**).
8.	Align $\hat{\mathbf{x}}$ and $\mathbf{x}^{(d)}$ using the standard Procrustes analysis.
9.	Apply the pose parameters from step (**8**) to the source shape $\mathbf{x}^{(s)}$.

3   Results

For numerical assessment, the robustness of the proposed technique is evaluated with respect to the severity of regional shape differences. To this end, a 3D liver dataset (Fig. 3) is used to synthetically create regional deformations, by randomly selecting and deforming a localized group of landmarks based on uniform noise. Varying percentages of deformed landmarks and amplitudes of deformation are simulated as shown in Fig. 2. Additionally, a perturbation of all landmarks is carried out using zero mean Gaussian noise. For comparison purposes, the standard Procrustes and a robust extension based on M-estimators [6] are implemented. It is evident from the results in Fig. 2 that the proposed technique outperforms existing Procrustes methods.

The improvement becomes particularly marked as the number of extreme residuals (a) and their amplitude increase (b). The robust Procrustes improves upon the least squares results up to a certain percentage of landmarks involved in the localized dissimilarity (around 20%). The proposed alignment, however, displays a higher breakdown point (close to 50%) and can handle large residuals as a result of the invariant influential point detection. Fig. 3 shows an example of regional shape deformation involving the left lobe of the liver (shown by the arrows). The influential points as detected by the proposed technique are displayed in dark shading in (c),

where it can be seen that the left lobe is correctly highlighted. This is a key feature of the proposed alignment which can be used as a quantitative as well as a visualization tool for shape interpretation.
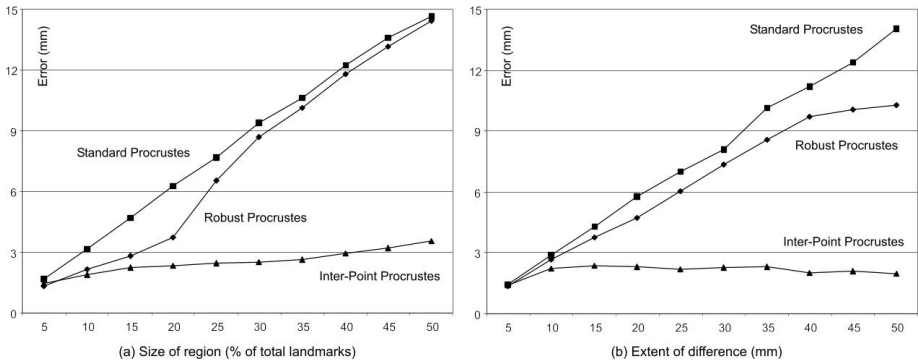
Fig. 2. Effects of the number of landmarks (a) and extent of localized deformation (b) on shape alignment by using the proposed algorithm and the techniques used for comparison. In (a), the amplitude is fixed to 30 mm, while in (b) the percentage of deformed landmarks equals 25%.
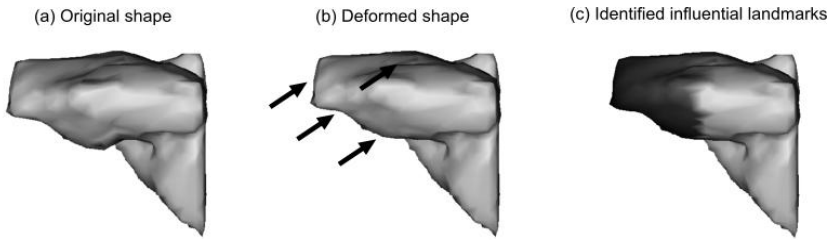
Fig. 3. Illustration of the performance of the proposed algorithm in explicit identification and visualization of regional shape differences

The proposed technique is then applied to a real dataset consisting of human carotid arteries. Such anatomical tree-like structures are very prone to large regional differences. In particular, existing research has shown that the geometry of the bifurcation is an important heamodynamic factor for the genesis and progression of atherosclerotic plaque. In this application, we have collected two classes of real carotid artery datasets. The first class contains 26 carotids with normal bifurcation geometry (example in Fig. 5(a)) and the second consists of three carotids with a genetic distortion of the external branch (example in Fig. 5(b)) as identified by the clinician (and referred to as wide-angle bifurcations). Both the standard and the inter-point Procrustes were then applied to the entire sample for shape discrimination and the results based on the first two modes of variation are plotted in Fig. 4. It can be seen that the separation between the two classes with the existing Procrustes is rather subtle, yet these shapes are significantly different as illustrated in Figs. 5(a) and (b). With the proposed technique, however, the class separation becomes pronounced.

This is because the proposed technique explicitly identifies the major regional differences (in this case the external branch), which are then taken into account during the intermediate alignment step. Unlike the method in [3], no supervision and user input is required to separate the three abnormal cases from the rest of the arteries.
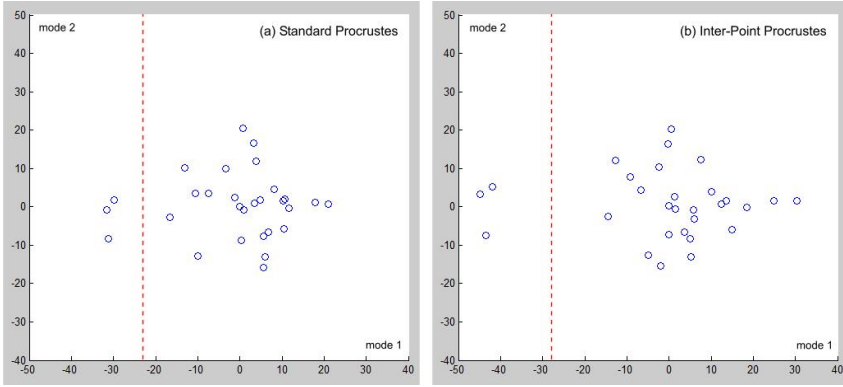
Fig. 4. Discrimination of the real carotid datasets by both the standard and proposed Procrustes
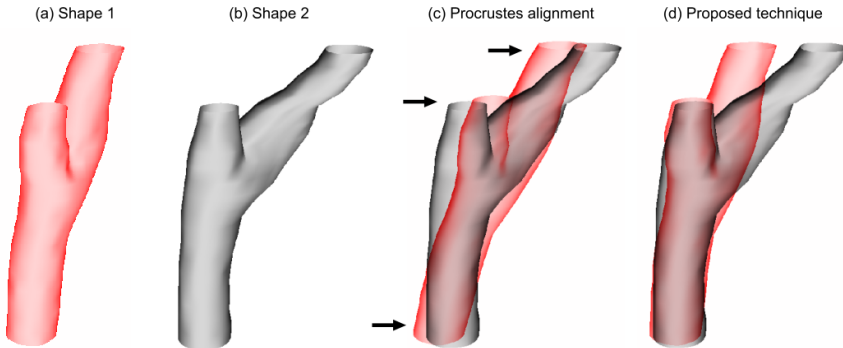
Fig. 5. Illustration of the strength of the proposed method for highlighting regional differences in tree-like structures. The carotid arteries in (a) and (b) differ significantly in their external branch, as detected by the proposed technique (d). The Procrustes analysis in (c), however, introduces alignment errors at various areas of the artery as shown by the arrows.

The strength of the proposed is further demonstrated in Fig. 5, where two carotids taken from the two classes are aligned with both the least-square Procrustes and the proposed method. Figs. 5(a) and 5(b) differ in various areas of the artery but with an evident large dissimilarity in the shape of their external branches (see Fig. 5(b)). In such a situation, the Procrustes analysis in Fig. 5(c) tends to distribute the residuals along the carotid, thus introducing alignment artifacts at the main branch and at the bifurcation as shown by the arrows. Although the gold standard alignment for real datasets is unknown a priori, it is evident that the proposed technique enables

improved fitting of the main and internal arteries, while the difference in the external artery is now fully highlighted. This demonstrates the benefit of the introduced technique to obtain anatomically meaningful results.

4   Conclusions

We have presented an inter-point Procrustes technique for robust shape alignment and interpretation. Unlike the standard method which minimizes a predefined matching criterion, the proposed approach is geometrically and biologically motivated: it explicitly identifies the common structures and the dissimilar regions in the shapes under investigation. This important information is taken into account during the alignment, which means the true shape differences can be highlighted in subsequent analysis. The validation shows that the technique can handle single-part (*e.g.*, liver) or multi-part (*e.g.*, tree-like structures) morphologies with a significant robustness to the severity of the regional differences. The results also suggest a potential interpretative value particularly for regional shape analysis and anatomical classification.

References

1. Dryden, I.L., Mardia, K.V.: Statistical shape analysis. Wiley, New York (1998)
2. Cootes, T.F., et al.: Active shape models - Their training and application. Computer Vision and Image Understanding (CVIU) 61(1), 38–59 (1995)
3. Loog, M., de Bruijne, M.: Discriminative Shape Alignment. In: Prince, J.L., Pham, D.L., Myers, K.J. (eds.) IPMI 2009. LNCS, vol. 5636, pp. 459–466. Springer, Heidelberg (2009)
4. Suinesiaputra, A., et al.: Automated detection of regional wall motion abnormalities based on a statistical model applied to multislice short-axis cardiac MR images. IEEE Transactions on Medical Imaging 28(4), 595–607 (2009)
5. Dorst, L.: First order error propagation of the procrustes method for 3D attitude estimation. IEEE Transactions on Pattern Analysis and Machine Intelligence 27, 221–230 (2005)
6. Verboon, P., Heiser, W.J.: Resistant orthogonal Procrustes analysis. Journal of Classification 9, 237–256 (1992)
7. Larsen, R.: L1 generalized Procrustes 2D shape alignment. Journal of Mathematical Imaging and Vision 31, 189–194 (2008)
8. Huber, P.J.: Robust statistics. Wiley, New York (1981)

Selection of Optimal Hyper-Parameters for Estimation of Uncertainty in MRI-TRUS Registration of the Prostate

Petter Risholm*, Firdaus Janoos, Jennifer Pursley, Andriy Fedorov,
Clare Tempany, Robert A. Cormack, and William M. Wells III

Brigham and Women's Hospital, Harvard Medical School
`pettri@bwh.harvard.edu`

Abstract. Transrectal ultrasound (TRUS) facilitates intra-treatment delineation of the prostate gland (PG) to guide insertion of brachytherapy seeds, but the prostate substructure and apex are not always visible which may make the seed placement sub-optimal. Based on an elastic model of the prostate created from MRI, where the prostate substructure and apex are clearly visible, we use a Bayesian approach to estimate the posterior distribution on deformations that aligns the pre-treatment MRI with intra-treatment TRUS. Without apex information in TRUS, the posterior prediction of the location of the prostate boundary, and the prostate apex boundary in particular, is mainly determined by the pseudo stiffness hyper-parameter of the prior distribution. We estimate the optimal value of the stiffness through likelihood maximization that is sensitive to the accuracy as well as the precision of the posterior prediction at the apex boundary. From a data-set of 10 pre- and intra-treatment prostate images with ground truth delineation of the total PG, 4 cases were used to establish an optimal stiffness hyper-parameter when 15% of the prostate delineation was removed to simulate lack of apex information in TRUS, while the remaining 6 cases were used to cross-validate the registration accuracy and uncertainty over the PG and in the apex.

1 Introduction

In conventional trans-rectal ultrasound (TRUS) guided prostate brachytherapy, low dosage radioactive seeds are permanently implanted throughout the prostate gland (PG). However, as the majority of prostate cancer cases are confined to the peripheral zone (PZ), partial gland implants targeting the PZ, which minimize the amount of healthy tissue irradiated, are desirable in treating prostate tumors. However, with TRUS it is difficult to determine the internal prostate substructure [1], and the boundary of the prostate apex (PA) (inferior part of prostate) is difficult to determine because of poor contrast from the constrained TRUS field of view and shadowing artifacts. The limited visibility of the gland in TRUS may lead to less than ideal placement of the radioactive seeds, resulting in iatrogenic complications. MR images obtained with an endorectal coil (ERC) have excellent soft tissue contrast, making them ideal for imaging the gland,

---

* Corresponding author.

its substructure and adjacent structures. Therefore, non-rigidly registering pre-treatment with the intra-treatment prostate images to enhance the delineation of prostate substructure is clinically important and shown to be feasible with intensity-based [2], model-based [1] and learning-based registration methods [3].

Registration methods typically optimize two terms – a likelihood (data) term and a weighted prior (regularization) term. At poorly visible substructures like the apex, the registration is driven more by the prior model and less by the data. Therefore, it becomes critical to optimize the free hyper-parameters (HPs) (*e.g.* regularization weight) of the registration model in order to minimize the error in their registered delineation [4]. Apart from manually tuning these hyper-parameters, one automated approach is to estimate an optimal value of the HP through minimization of cross-validation error in localizing homologous points in the registered fixed and moving images [4]. However, this approach only uses point-estimates of the registration parameters, and discards the associated uncertainty that a posterior distribution provides, when determining the optimal HP setting.

In contrast, Bayesian registration frameworks [5,6,7] that quantify the posterior distribution on transformation parameters can estimate the uncertainty in registration in terms of credibility intervals or inter-quartile ranges, along with an optimal alignment. For example, Risholm *et al.* [6] estimate the posterior distribution $p(\mathbf{U} \mid \mathcal{P}_{us}, \mathcal{P}_{mr}, \tau)$ of the transformation $\mathbf{U}$ using a finite-element model [6] which treats the prostate as a synthetic elastic bio-mechanical object. Here, $\mathcal{P}_{us}$ and $\mathcal{P}_{mr}$ are representations of the prostate in TRUS and MR images respectively which are demarcated as part of the standard brachytherapy workflow. Furthermore, assume that $\mathcal{P}_{mr}$ and $\mathcal{P}'_{us} = \mathcal{P}_{us} \bigcup \mathcal{A}_{us}$ are the ground truth representations of the prostate, that $\mathcal{A}_{us}$ represents the missing apex in the TRUS, and that the homologous area of $\mathcal{A}_{us}$ in the MR image is represented as $\mathcal{A}_{mr} \subset \mathcal{P}_{mr}$.

Boltzmann's distribution is used to convert the elastic energy of a deformation into a prior $p(\mathbf{U} \mid \tau)$, where τ is a temperature HP that can be interpreted as an overall inverse pseudo stiffness of the underlying tissue. The value of this temperature HP affects both the spread/uncertainty of the distribution as well as the mode, especially in areas with weak data like the apex. One strategy [5,8] for dealing with HPs in a Bayesian setting is to equip them with a maximum entropy hyper-prior $p(\tau)$ and marginalize them out from the posterior distribution: $p(\mathbf{U} \mid \mathcal{P}_{us}, \mathcal{P}_{mr}) = \int p(\tau) p(\mathbf{U} \mid \mathcal{P}_{us}, \mathcal{P}_{mr}, \tau) d\tau$. However, the drawback with marginalization under an over-dispersed hyper-prior is that it may result in an over-dispersed posterior, as compared to a hyper-prior concentrated around an appropriate value of the HP. This effectively implies increased parameter uncertainty due to HP uncertainty, and this effect will be strong in areas with weak data (*e.g.* $\mathcal{A}_{us}$).

The main contribution of this paper is a technique for HP selection under uncertainty, where we maximize the likelihood of the temperature τ under the posterior predictive distribution $p(\partial \mathcal{A}_{us} \mid \partial \mathcal{A}_{mr}, \tau)$ of the apex boundary $\partial \mathcal{A}_{us}$ in the TRUS image conditioned on its boundary $\partial \mathcal{A}_{mr}$ in the MR image. This

distribution is obtained from the posterior distribution of the transformation $p(\mathbf{U} \mid \mathcal{P}_{\mathrm{us}}, \mathcal{P}_{\mathrm{mr}}, \tau)$. Consequently, our method identifies a specific value of the temperature HP that maximizes the ability of the posterior distribution in predicting poorly determined structures like the apex while mitigating the effect of HP related uncertainty. The proposed method for estimating optimal HPs is not restricted to MRI-TRUS registration, but is applicable in any situation where we have ground truth homologous correspondences in the two image spaces.

Because the apex is difficult to delineate in TRUS, the method was trained and validated on a data-set of 10 corresponding pre- and intra-treatment prostate MRI, where accurate ground truth delineation of the prostate gland boundary was possible. With the intra-treatment MRI acting as a surrogate for the TRUS, the apex was systematically removed so that the prior model was the main predictor of the apex location. Temperature estimation was carried out on 4 cases, while the remaining 6 cases were used to cross validate the model's ability to predict the location of the apex. These results are presented in Section 3 and are representative for MRI-TRUS registration results assuming that the deformations seen in the pseudo TRUS (*e.g.* intra-treatment MRI) are good approximations of the deformations seen in real TRUS images.

2   Methods and Materials

2.1   Bio-mechanical Prostate Model

As in Risholm *et al.* [6], we model the prostate with a bio-mechanical finite-element (FE) mesh that assumes the prostate $\mathcal{P}_{\mathrm{mr}}$ to be an elastic material with unknown stiffness and compressibility parameters. The tetrahedral FE-mesh consists of N_e tetrahedra and N_v vertices $\mathbf{V} = [\mathbf{v}_1^\top, \ldots, \mathbf{v}_{N_v}^\top]$. The boundary vertices $\mathbf{v}_i = [x_i, y_i, z_i]^\top$ are identified as $\mathbf{V}_b = [\mathbf{v}_1^\top, \ldots, \mathbf{v}_{N_b}^\top]$ with $\mathbf{v}_i \in \partial\mathcal{P}_{\mathrm{mr}}$, where $N_b \leq N_v$. Each boundary vertex is associated with a displacement vector $\mathbf{u}_i = [u_i, v_i, w_i]^\top$ such that $\mathbf{v}_i[u_i] = \mathbf{v}_i + \mathbf{u}_i \in \partial\mathcal{P}_{\mathrm{us}}$. Once the deformation of the boundary $\mathbf{U} = [\mathbf{u}_1^\top, \ldots, \mathbf{u}_{N_b}^\top]^\top$ that aligns $\partial\mathcal{P}_{\mathrm{mr}}$ with $\partial\mathcal{P}_{\mathrm{us}}$ is determined, the displacement of the internal nodes under the linear elastic model is determined by solving a linear system of equations. Associated with each pair of vertices $\mathbf{v}_i$ and $\mathbf{v}_j$ is a stiffness matrix $\mathbf{K}_{ij}$ which depends on Young's modulus ($E = 1/\tau$) and Poisson's ratio (ρ) of the underlying tissue. Because no boundary forces are applied, E is a unit-less pseudo stiffness. The linear elastic (band) energy for the deformation of the prostate boundary is defined as $E_{el}(\mathbf{U}) = \sum_{i=1}^{N_b} \sum_{j \in \mathcal{B}_i} \mathbf{u}_i \mathbf{K}_{ij} \mathbf{u}_j$, where $\mathcal{B}_i$ is the set of vertices in the Markov neighborhood of vertex i.

2.2   Bayesian Estimation of Boundary Conditions

Under the posterior distribution $p(\mathbf{U} \mid \mathcal{P}_{\mathrm{us}}, \mathcal{P}_{\mathrm{mr}}, \tau) \propto p(\mathcal{P}_{\mathrm{us}} \mid \mathbf{U}, \mathcal{P}_{\mathrm{mr}})p(\mathbf{U} \mid \tau)$, the likelihood term corresponds to the distance of the deformed boundary $\partial\mathcal{P}_{\mathrm{mr}} \circ \mathbf{U}$ from the TRUS boundary $\partial\mathcal{P}_{\mathrm{us}}$, and is a normal distribution on the ℓ_2 distance from $\partial\mathcal{P}_{\mathrm{us}}$:

$$p(\mathcal{P}_{\mathrm{us}} \mid \mathbf{U}, \mathcal{P}_{\mathrm{mr}}) = \mathcal{N}(||\partial\mathcal{P}_{\mathrm{mr}} \circ \mathbf{U} - \partial\mathcal{P}_{\mathrm{us}}||_2^2 \mid 0, \sigma^2) , \tag{1}$$

where σ^2 models the uncertainty of the position of the boundary $\partial\mathcal{P}_{\mathrm{us}}$.

The prior distribution on deformations uses Boltzmann's distribution on the linear elastic energy $E_{el}(\mathbf{U})$ parametrized by the temperature τ so the log posterior becomes

$$\log p(\mathbf{U} \mid \mathcal{P}_{\mathrm{us}}, \mathcal{P}_{\mathrm{mr}}, \tau) = -\frac{1}{\sigma^2} \sum_{i=1}^{N_b} \triangle^2(\mathbf{v}_i[u_i]; \partial\mathcal{P}_{\mathrm{us}}) - \frac{E_{el}(\mathbf{U})}{\tau} + \mathrm{const}, \qquad (2)$$

where $\triangle$ is the (signed) ℓ_2 distance of vertex $\mathbf{v}_i[u_i]$ from $\partial\mathcal{P}_{\mathrm{us}}$.

The Boltzmann's temperature controls the overall stiffness of the elastic material (*i.e.* it uniformly modulates all stiffness matrices $\mathbf{K}_{ij}$) and is, in general, a priori unknown. Higher temperatures correspond to lower overall stiffness, increases the spread of both the prior and hence the posterior distribution, while lower temperatures indicate stiffer materials with tight spreads of the distributions. Furthermore, this temperature also affects the location of the mode of the distributions.

The posterior distribution is characterized non-parametrically using Metropolis Hastings (MH) Markov Chain Monte Carlo (MCMC) sampling. Starting from an initial deformation $\mathbf{U}_0$, a proposal boundary deformation $\mathbf{U}_*$ is generated from a multi-variate normal proposal distribution and accepted ($\mathbf{U}_n = \mathbf{U}_*$) or rejected ($\mathbf{U}_n = \mathbf{U}_{n-1}$) according to the MH-criterion. After convergence of the MCMC chain, assessed with the Geweke criterion, the collection of posterior samples $\mathcal{U} = \{\mathbf{U}_1, \ldots, \mathbf{U}_{N_s}\}$ characterizes the posterior distribution in Eqn. (2).

2.3   Temperature Estimation

This section describes how to estimate the temperature HP through maximization of the posterior predictive distribution given a representative training data-set with ground truth markers, either in the form of homologous points, contours or surfaces. In this paper, these markers are the manually demarcated boundaries of the prostate apex, $\partial\mathcal{A}_{\mathrm{us}} \subset \partial\mathcal{P}'_{\mathrm{us}}$ and $\partial\mathcal{A}_{\mathrm{mr}} \subset \partial\mathcal{P}_{\mathrm{mr}}$, extracted from pre- and intra-treatment prostate MRI, where the intra-treatment MRI is used as a surrogate for TRUS. Note that the apex boundary $\partial\mathcal{A}_{\mathrm{us}}$ is not used to estimate the posterior distribution on deformations, but is used only post-registration in the temperature estimation step. Therefore, the distribution of the location of the apex boundary $\partial\mathcal{A}_{\mathrm{mr}}^{\mathrm{us}}$ as *predicted by* the MR image is:

$$p(\partial\mathcal{A}_{\mathrm{mr}}^{\mathrm{us}} \mid \partial\mathcal{A}_{\mathrm{mr}}, \tau) = \int p(\partial\mathcal{A}_{\mathrm{mr}}^{\mathrm{us}} \mid \mathbf{U}, \partial\mathcal{A}_{\mathrm{mr}}) p(\mathbf{U} \mid \mathcal{P}_{\mathrm{us}}, \mathcal{P}_{\mathrm{mr}}, \tau) d\mathbf{U} , \qquad (3)$$

where $p(\partial\mathcal{A}_{\mathrm{mr}}^{\mathrm{us}} \mid \mathbf{U}, \partial\mathcal{A}_{\mathrm{mr}}) = \delta\left(\|\partial\mathcal{A}_{\mathrm{mr}} \circ \mathbf{U} - \partial\mathcal{A}_{\mathrm{mr}}^{\mathrm{us}}\|_2^2\right)$ is modeled with a Dirac delta which is zero everywhere except where, after applying a transformation $\mathbf{U}$, the apex boundary from the MR image $\partial\mathcal{A}_{\mathrm{mr}}$ overlaps with the candidate apex boundary $\partial\mathcal{A}_{\mathrm{mr}}^{\mathrm{us}}$. The posterior probability of a particular deformation $p(\mathbf{U} \mid \mathcal{P}_{\mathrm{us}}, \mathcal{P}_{\mathrm{mr}}, \tau)$ is given by the set of deformation samples generated through MCMC (§Section 2.2) using non-parametric Kernel Density Estimation (KDE).

Therefore, integrating out $\mathbf{U}$ in Eqn. (3), and through the sifting property of the Dirac delta function, the kernel estimate of predictive density becomes:

$$p(\partial \mathcal{A}_{\mathrm{mr}}^{\mathrm{us}} \mid \partial \mathcal{A}_{\mathrm{mr}}, \tau) = \frac{1}{N_s} \sum_{s=1}^{N_s} k_h \left(||\partial \mathcal{A}_{\mathrm{mr}} \circ \mathbf{U}_s - \partial \mathcal{A}_{\mathrm{mr}}^{\mathrm{us}}||_2^2 \right) , \qquad (4)$$

where k is a symmetric positive kernel with bandwidth h, mean 0 and integral 1, and $\mathbf{U}_s$ is the s–th sample from the MCMC chain of N_s samples. Now, given a training data-set of D_t pairs of ground-truth prostate delineations from pre- and intra-treatment images, and assuming independence conditioned on the temperature, the *maximum likelihood estimate* (MLE) of the temperature HP from the posterior predictive distribution is:

$$\tau^* = \operatorname*{argmax}_{\tau} \prod_{d=1}^{D_t} p(\mathcal{A}_{\mathrm{mr}}^{\mathrm{us}} = \mathcal{A}_{\mathrm{us}}^{(d)} \mid \mathcal{A}_{\mathrm{mr}}^{(d)}, \tau), \qquad (5)$$

where $\mathcal{A}_{\mathrm{us}}^{(d)}$ is the manually demarcated boundary of the apex in the surrogate TRUS image. Hence, we choose the model temperature τ such that the ground truth position of the apex has highest predictive probability.

2.4   Patient Data

In this evaluation, intra-treatment MRI was used as a proxy for TRUS because it permitted an accurate demarcation of the apex boundary. The patient data, collected from $D = 10$ patients undergoing prostate biopsy, was acquired as part of a prospective clinical research protocol which is HIPPA compliant and approved by the local institutional review board. For each patient, a diagnostic MRI ($512 \times 512 \times 30$ with a pixel size of $0.3 \times 0.3 \times 3$ mm^3) and an intra-treatment MRI ($320 \times 320 \times 40$ with pixel size $0.5 \times 0.5 \times 3$ mm^3) used for biopsy guidance were acquired. There is significant change in the prostate between the two images due to the use of an ERC when acquiring the diagnostic MRI and different patient positioning during imaging. The PGs were segmented for each patient on both the pre- and intra-treatment MR images by an abdominal radiologist with over 10 years of experience in prostate MR interpretation.

3   Results

The label-maps of the pre- and intra-treatment segmentations were rigidly registered, and the manually contoured boundaries were smoothed through anti-aliasing followed by Gaussian smoothing with variance of 1mm^2. For each of the D cases, a tetrahedral FE-mesh of $\mathcal{P}_{\mathrm{mr}}^{(d)}$ was created. We also removed the apex $\mathcal{A}_{\mathrm{us}}^{(d)}$, corresponding to approximately 15% of the prostate volume, from the ground truth label-map of the surrogate TRUS image such that $\mathcal{P}_{\mathrm{us}}^{(d)} = \mathcal{P}_{\mathrm{us}}^{(d)'} \backslash \mathcal{A}_{\mathrm{us}}^{(d)}$, in order to evaluate the ability of the model to correctly predict the location of the apex in the total absence of image contrast.

The likelihood variance was set to the maximum pixel spacing (3mm) to model the boundary label uncertainty, and the prostate was modeled as a nearly incompressible material ($\rho = 0.45$). The average number of FE-vertices over the 10 cases was $\bar{N}_v = 415$. In all MCMC runs, 2×10^6 samples were generated, which after thinning with a factor of 10 and discarding the first 25% of the chain for burn-in, resulted in $N_s = 1.5 \times 10^5$ remaining samples. The total running time for one chain was ≈ 4 hrs.

3.1 Temperature Estimation

We first performed MLE (§Section 2.3) for the temperature HP on $D_t = 4$ randomly chosen data-sets. To optimize Eqn. (3), the temperature τ was discretized into 10 bins over the range $20 \leq \tau \leq 380$ and the predictive density (§Eqn. (4)) was estimated for each temperature and data-set in parallel (4×10 parallel computations). Because of the high number of samples, the particular choice of kernel and bandwidth in Eqn. (4) has a negligible effect on the result – we used a standard normal kernel and found the optimal temperature to be $\tau^* = 180$. Figure 1 shows the marginal posterior distribution of the deformed prostate in relation to the ground truth at three different temperature settings and shows that uncertainty plays an important role in selecting HPs.

(a) $\tau = 20$ (b) $\tau = 180$ (c) $\tau = 340$
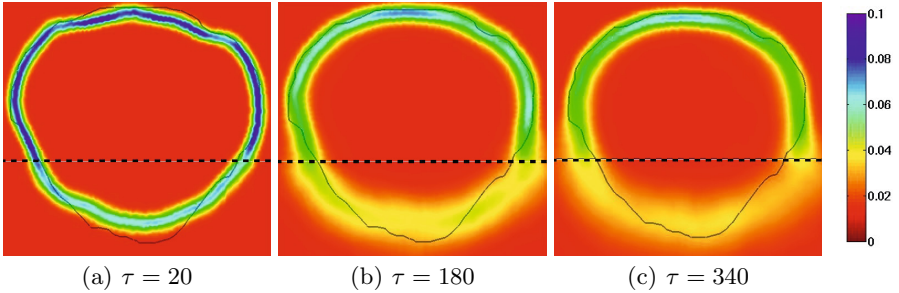
Fig. 1. Coronal slices of the marginal posterior probability $p(\partial\mathcal{P}_{\mathrm{mr}}^{\mathrm{us}} \mid \phi, \theta)$ of $\partial\mathcal{P}_{\mathrm{mr}}^{\mathrm{us}}$ crossing a ray, with spherical coordinates ϕ and θ, originating from the prostate center at three different temperatures for one of the 4 training cases. The region below the horizontal line denotes $\mathcal{A}_{\mathrm{us}}$, and the black outline delineates $\partial\mathcal{P}_{\mathrm{us}}'$. For each temperature we compare the predicted log probability $\Xi(\tau) = \log p(\partial\mathcal{A}_{\mathrm{us}} \mid \partial\mathcal{A}_{\mathrm{mr}}, \tau)$ with the conventional Expected Mean Squared Error (EMSE): $\Lambda(\tau) = \frac{1}{N_b} \sum_{i=1}^{N_b} \int \triangle^2 (\mathbf{v}_{u_i}, \partial\mathcal{A}_{\mathrm{us}}) \, p(\mathbf{U} \mid \mathcal{P}_{\mathrm{us}}, \mathcal{P}_{\mathrm{mr}}, \tau) d\mathbf{U}$. **Fig. (a):** $\Xi(20) \approx -\infty$ and $\Lambda(20) = 22.7$. **Fig. (b):** $\Xi(180) = -69.8$ and $\Lambda(180) = 37.9$. **Fig. (c):** $\Xi(340) = -73.1$ and $\Lambda(340) = 66.3$. The trend, which is confirmed by the 3 other training data-sets, is that EMSE is overly optimistic about distributions with low temperatures, and that these distributions, as shown in Fig. (a), are implausible when evaluated on the training data. Conversely, the MLE chooses the temperature ($\tau^* = 180$) that maximizes the posterior predictive distribution on the training data.

3.2 Evaluation of Method

The accuracy and precision of the HP estimation method was evaluated using the 6 remaining cases. The predictive distribution of both the apex boundaries and the remaining prostate gland were characterized for each case at τ^*, and error in boundary prediction as well as the inter-quartile range (IQR) on the distance errors were computed, as listed in Table 1. Figure 2 shows the marginal posterior distribution of the prostate, at optimum temperature, which conveys the mode as well as the corresponding uncertainty and is compatible with the ground-truth delineation of the prostate.
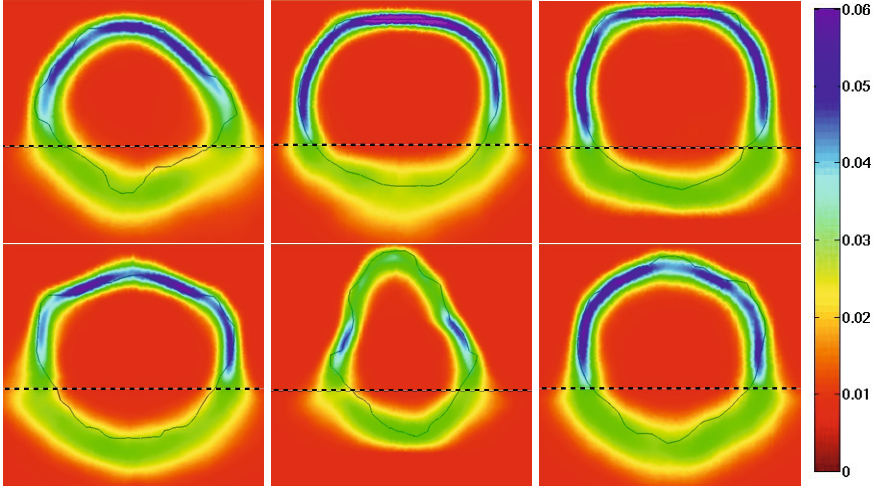
Fig. 2. Results from the evaluation using the optimal temperature ($\tau^* = 180$). Coronal slices of the marginal posterior probability $p(\partial\mathcal{P}_{\mathrm{mr}}^{\mathrm{us}} \mid \phi, \theta)$ of $\partial\mathcal{P}_{\mathrm{mr}}^{\mathrm{us}}$ crossing a ray, with spherical coordinates ϕ and θ, originating from the prostate center. The ground truth boundary of the prostate $\partial\mathcal{P}_{\mathrm{us}}'$ is overlaid as a black contour. The area below the horizontal line delineates $\mathcal{A}_{\mathrm{us}}$.

Table 1. Accuracy and precision of the posterior distribution at optimal temperature value for PA ($\partial\mathcal{A}_{\mathrm{mr}}^{\mathrm{us}}$) and the rest of the PG ($\partial\mathcal{P}_{\mathrm{mr}}^{\mathrm{us}} \setminus \partial\mathcal{A}_{\mathrm{mr}}^{\mathrm{us}}$). **Note:** All values in mm. For each case we estimated the mode $\hat{\mathbf{u}}$ of the posterior distribution. The mean and maximum errors are the mean and maximum absolute distances of boundary vertices deformed by $\hat{\mathbf{u}}$ from $\partial\mathcal{P}_{\mathrm{us}}'$ measured by the $\triangle$ metric. Also tabulated is the inter-quartile range (IQR) of the distribution of the distances $\{\triangle(\mathbf{v}_i \circ \mathbf{u}_i^s; \partial\mathcal{P}_{\mathrm{us}}')\}_{i \in \partial\mathcal{P}_{\mathrm{mr}}, s=1,\ldots N_s}$.

	Case:	1	2	3	4	5	6
	Mean Error	3.9	5.5	5.0	4.7	4.7	4.5
Apex	Max Error	8.5	12.9	11.6	12.3	10.7	15.4
	IQR	6.6	8.1	7.4	7.4	7.6	8.8
	Mean Error	1.6	1.7	1.8	1.7	1.9	1.6
Prostate	Max Error	14.1	12.7	11.5	15.0	12.9	11.4
	IQR	3.8	4.3	4.6	4.1	5.4	3.6

4   Discussion

With a Bayesian method, we estimate the posterior distribution on boundary conditions of an elastic model of the prostate extracted from MRI that will align it with the prostate delineated in TRUS. Because the posterior distribution, and in particular the posterior prediction of the location of the prostate apex, is sensitive to the pseudo stiffness (prior) HP, we maximize the posterior predictive distribution with regards to the HP given homologous ground truth apex delineation. The registration method, with optimal HPs, was evaluated on 6 MRI-MRI datasets where the mean error on the boundary was found to be 1.7mm. Comparable results are reported on homologous landmarks, *e.g.* using intensity-based registration [2] with a mean error of 1.5mm and the model-based approach of Hu *et al.* [1] with median RMS of 2.4mm. However, the advantage of our method is that we prescribe error-bars on the registration results, in terms of IQRs, and that the true prostate boundary is within these error-bars.

References

1. Hu, Y., Ahmed, H.U., Taylor, Z., Allen, C., Emberton, M., Hawkes, D., Barratt, D.: MR to ultrasound registration for image-guided prostate interventions. Med. Image Anal. 16(3), 687–703 (2012)
2. Fedorov, A., Tuncali, K., Fennessy, F.M., Tokuda, J., Hata, N., Wells, W.M., Kikinis, R., Tempany, C.M.: Image registration for targeted MRI-guided transperineal prostate biopsy. J. Magn. Reson. Imaging (2012)
3. Shi, Y., Liao, S., Shen, D.: Learning statistical correlation for fast prostate registration in image-guided radiotherapy. Med. Phys. 38(11), 5980–5991 (2011)
4. Yeo, B.T.T., Sabuncu, M.R., Vercauteren, T., Holt, D.J., Amunts, K., Zilles, K., Golland, P., Fischl, B.: Learning task-optimal registration cost functions for localizing cytoarchitecture and function in the cerebral cortex. IEEE Trans. Med. Imaging 29(7), 1424–1441 (2010)
5. Simpson, I.J., Schnabel, J.A., Groves, A.R., Andersson, J.L., Woolrich, M.W.: Probabilistic inference of regularisation in non-rigid registration. NeuroImage 59(3), 2438–2451 (2012)
6. Risholm, P., Fedorov, A., Pursley, J., Tuncali, K., Cormack, R., Wells, W.M.: Probabilistic non-rigid registration of prostate images: Modeling and quantifying uncertainty. In: IEEE Int. Symp. Biomed. Imaging, pp. 553–556 (2011)
7. Gee, J.C., Bajcsy, R.K.: Elastic matching: Continuum mechanical and probabilistic analysis. In: Brain Warping, pp. 183–197. Academic Press (1999)
8. Janoos, F., Risholm, P., Wells, W.M.: Robust non-rigid registration and characterization of uncertainty. In: Zhou, K., Duncan, J.S., Ourselin, S. (eds.) IEEE Workshop on Mathematical Methods in Biomedical Image Analysis, vol. 1, pp. 4321–4326 (2012)

Globally Optimal Deformable Registration on a Minimum Spanning Tree Using Dense Displacement Sampling

Mattias P. Heinrich[1,2,*], Mark Jenkinson[2], Sir Michael Brady[3],
and Julia A. Schnabel[1]

[1] Institute of Biomedical Engineering, University of Oxford, UK
[2] Oxford University Centre for Functional MRI of the Brain, UK
[3] Department of Oncology, University of Oxford, UK
mattias.heinrich@eng.ox.ac.uk,
http://users.ox.ac.uk/~shil3388

Abstract. Deformable image registration poses a highly non-convex optimisation problem. Conventionally, medical image registration techniques rely on continuous optimisation, which is prone to local minima. Recent advances in the mathematics and new programming methods enable these disadvantages to be overcome using discrete optimisation. In this paper, we present a new technique **deeds**, which employs a discrete **de**nse **d**isplacement **s**ampling for the deformable registration of high resolution CT volumes. The image grid is represented as a minimum spanning tree. Given these constraints a global optimum of the cost function can be found efficiently using dynamic programming, which enforces the smoothness of the deformations. Experimental results demonstrate the advantages of **deeds**: the registration error for the challenging registration of inhale and exhale pulmonary CT scans is significantly lower than for two state-of-the-art registration techniques, especially in the presence of large deformations and sliding motion at lung surfaces.

1 Introduction

Deformable image registration is a key enabling technique in medical image analysis. Applications include motion correction, image-guided radiotherapy, multi-modality fusion, quantification of longitudinal progression of disease, and inter-subject registration for atlas-based segmentation. Non-rigid registration algorithms aim to solve a highly non-convex optimisation problem with several million degrees of freedom. In the medical domain, this problem has so far been addressed almost exclusively using continuous optimisation. In a recent comparison study of non-rigid registration methods applied to pulmonary CT scans [1], 23 out of 24 algorithms used continuous optimisation. However, continuous optimisation of a non-convex cost function is susceptible to local minima, potentially leading to an erroneous registration. Local minima are frequently encountered

---

when the initial motion estimate is far from the desired solution, especially for small anatomical features, which have undergone large deformations. A second drawback for continuous optimisation is the necessity either for an analytical or for a numerical derivative of the cost function, thus limiting its choice.

Discrete optimisation offers numerous advantages to overcome these limitations. It has been widely used for 2D computer vision applications. However, in discrete optimisation the space of possible displacements $\mathcal{L}$ has to be quantised. If the smallest discretisation step is defined to be one voxel, the displacement space can be of the order of thousands and therefore there are millions of degrees of freedoms. This high computational and memory demands have previously discouraged its use in 3D problems. One state-of-the-art medical image registration method, based on discrete optimisation, **drop** [2], reduces the dimensionality of the problem by using a parametric transformation model based on a B-spline deformation grid. Additionally, the displacement space is sampled only sparsely (along the three axes). This may possibly cause the optimal displacement to be missed. The authors attempt to address this problem, first by introducing a multi-resolution scheme, and second by iteratively updating the transformation (thus warping the source image towards the target). While the method is an improvement (both in terms of accuracy and speed) over commonly used continuous optimisation counterparts, it lacks some of the attractive properties that a discrete framework potentially offers: most notably, avoidance of an iterative solution and image interpolation; and the use of a dense sampling of the displacement space.

In this work, we address these shortcomings in three ways. First, we reformulate the image grid so that it is not fully connected. Instead, a minimum spanning tree [3] is computed, which best replicates the underlying structure of the anatomical connectivity of the image. This allows us to use dynamic programming to find the global optimum of the registration cost non-iteratively in just two passes. Second, we use the min-convolution technique [4] to reduce the complexity of the pair-wise regularisation cost computation from $\mathcal{O}(|\mathcal{L}|^2)$ to $\mathcal{O}(|\mathcal{L}|)$. Finally, we use a multi-level approach in which groups of voxels are represented by a single node in the graph. This leads to an efficient coarse-to-fine optimisation, while, at the same time, all data terms are calculated in the original image resolution (so there is no image degradation as a result of downsampling). At subsequent (finer) levels, the previous solution is used as a prior of the displacement. The improved computational efficiency allows us to use a very large label space and thus better address the efficiency-quality trade-off than previous approaches. We employ a symmetric diffeomorphic approach to ensure unbiased physically plausible transformations. Our approach is called **deeds** (**de**nse **d**isplacement **s**ampling) and is explained in detail in the next section. An experimental validation is performed for the non-rigid registration of pulmonary CT scans at different breathing states. We compare our approach to a continuous optimisation algorithm (**gsyn** [5]), which performed best in the recent EMPIRE study [1]; and a discrete parametric optimisation framework (**drop**) using linear programming [2].

2   Deformable Registration Using deeds

Discrete optimisation is usually performed as Markov Random Field (MRF) labelling. For the purposes of our non-parametric image registration framework, a graph is defined, in which the nodes $p \in \mathcal{P}$ correspond to voxels (or groups of voxels) and in which, for each node, there is a set of (hidden) labels f_p, which correspond to discrete displacements. The energy function to be optimised consists of two terms: the data (also called unary) cost D (which is independent for each node); and the pair-wise regularisation cost $R(f_p, f_q)$ for any node q, which is directly connected ($\in \mathcal{N}$) with p:

$$E(f) = \underbrace{\sum_{p \in \mathcal{P}} D(f_p)}_{\text{data term}} + \alpha \underbrace{\sum_{(p,q) \in \mathcal{N}} R(f_p, f_q)}_{\text{regularisation term}} \tag{1}$$

The unary cost measures the similarity of a voxel in one image and a displaced voxel in the second image, and is independent of the displacements of its neighbours. The pair-wise term enforces a globally smooth transformation by penalising deviations of the displacements of neighbouring voxels. The weighting parameter α sets the influence of the regularisation.

Methods to solve the MRF labelling problem can generally be categorised as one of two approaches: message passing and graph cuts. Message passing schemes include: loopy belief propagation (LBP); sequential tree-reweighted message passing (TRW-S); and dynamic programming on a tree. Popular graph cut algorithms include: α-expansion moves; and the fast primal-dual strategy (FastPD). For some problems, graph cuts guarantee convergence to a known bound, close to the global optimum. However, in practice, the complexity limits both the number of nodes and the label space (directly applied to medical images α-expansion graph cuts take up to 24 hours for a single registration [6]). In contrast, our approach can find the global optimum for a complex 3D registration problem with a very large (dense) label space within minutes using a reduced neighbourhood interaction based on a minimum spanning tree (MST).

2.1   Dynamic Programming on Minimum Spanning Tree (MST)

Optimising the registration cost function on a six-connected graph is NP-hard. In contrast, the very efficient dynamic programming technique finds a global minimum, without iterations, in just two passes on a cycle-free graph (e.g. a tree). Using Prim's algorithm [7], we can quickly find the unique MST given a set of nodes $p \in \mathcal{P}$ and edges e. The edge weight $w(p, q)$ is defined as the sum of absolute differences (SAD) between the intensities of all voxels contained within the node p and the respective voxels in node q. The MST is a spanning tree with minimum total edge costs. The selection of the root node does not influence the result of the optimisation. The MST sufficiently reflects the underlying anatomical connectivity in a medical image (see Fig. 1 left). It is well balanced, and, as a consequence, the maximum width is approximately $|\mathcal{P}| / \log|\mathcal{P}|$. The output
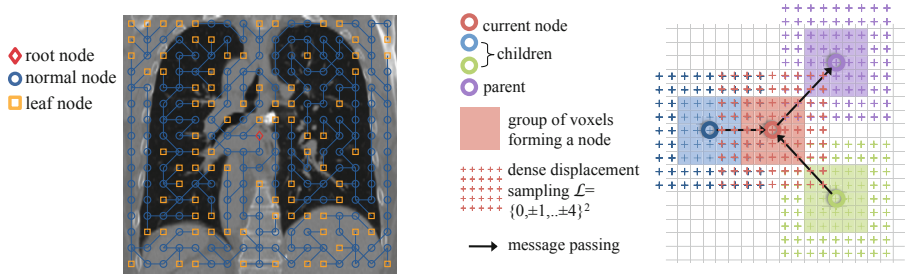
Fig. 1. An example of minimum spanning tree (MST) of a 2D coronal slice of a lung CT is shown on the left ($\diamond$ root, $\bigcirc$ normal, and $\square$ leaf nodes). The concept of our proposed discrete optimisation scheme, using dense displacement sampling, is displayed on the right. Note that, even though the nodes may be sparsely distributed, the data cost is always computed in the original resolution.

of Prim's algorithm consists of a sorted list of all nodes (with increasing tree depth) and the index of each node's parent. A similar approach has been used for stereo correspondence [3] however, other methods (LBP, TRW-S) perform better on that specific application.

Finding the best labelling, and thereby the global optimum of Eq. 1, is possible using dynamic programming on the MST [3]. At each node p, the cost C_p of the best displacement can be found, given the displacement f_q of its parent q:

$$C_p(f_q) = \min_{f_p} \left(D(f_p) + R(f_p, f_q) + \sum_c C_c(f_p) \right) \tag{2}$$

where c are the children of p. The best displacement can be found by replacing min with argmin in Eq. 2. For any leaf node, Eq. 2 can be evaluated directly (since it has no children). Thereafter, the tree is traversed from its leaves down to the root node. It is worth noting that the costs C_p only have to be stored for the next tree level (only the argmin is needed to select the best displacement). Once the root node is reached, the best labelling for all nodes can be selected in another pass through the tree (from root to leaves). Finding the minimum naïvely would require $|\mathcal{L}|^2$ calculations for the regularisation cost per node. In [4] the min-convolution technique is introduced, which reduces the complexity to $|\mathcal{L}|$ by employing a lower envelope computation. For most commonly used (pair-wise) regularisation terms, such as diffusion (squared difference of displacements) and total variation (absolute difference) regularisation, this simplification is possible.

2.2 Dense Displacement Sampling

To avoid local minima, most continuous-optimisation-based registration algorithms use a multi-resolution scheme in which the images are downsampled after a prior Gaussian smoothing. This may degrade the quality of the registration. We adopt a different approach: a multi-level scheme, in which we only employ the highest resolution image. For a given level, the image is subdivided into

non-overlapping cubic groups of voxels. The similarity cost is first calculated for each voxel separately using dense displacement sampling, then aggregated for all voxels of the same group (this forms an additional intrinsic regularisation and reduces the number of nodes). Subsequently, the regularisation term is calculated only for each group of voxels (see Fig. 1). Using this approach, both high spatial accuracy and low computational complexity are achieved. The optimal labelling is obtained as explained previously. For the next level, the previous deformation field is upsampled and used as the prior deformation. At a finer level, the sampling range may be reduced, because the optimal transformation for the coarser level is known and only a small deviation from it is expected.

2.3   Symmetric and Diffeomorphic Transformations

For many deformable registration algorithms, one image has to be chosen as the target, the other as the moving image. This biases the registration outcome and may additionally introduce an inverse consistency error (ICE). The ICE for a forward transform A and a backward transform B is defined as the difference between AB and the identity. **gsyn** [5] uses a symmetric deformable registration, which calculates a transformation from both images to a common intermediate image and also ensures that $A(0.5) = B(0.5)^{-1}$. The full forward transformation is then $A(0.5) \circ B(0.5)^{-1}$. We adopt a similar approach and estimate both A and B. We then use a fast iterative inversion method, as presented in [8], to obtain $A(0.5)^{-1}$ and $B(0.5)^{-1}$. Additionally, if we treat the displacement field as a velocity field, we can transform it into a diffeomorphic mapping by applying the scaling and squaring method [9]. This approach avoids transformations for which physically implausible folding of volume occurs. Although this yields a continuous-valued transformation, we need to apply the voxel-sized discretisation for the next level. However, the discretisation is not performed for the final level.

3   Experiments and Results

We evaluate our deformable registration method on ten cancer (esophagus or lung) patients of the DIR-Lab 4D CT dataset[1] between the inhale and exhale phases (no intermediate frames were used in our experiments). The scans were acquired as a breathing cycle CT of the thorax and upper abdomen with a spatial resolution of 0.97 to 1.16 mm in the xy-direction and 2.5 mm in the z-direction. These registration tasks are particularly challenging due to: the large deformations of small features (lung vessels, airways); the discontinuous sliding motion between lung lobes and the lung rib cage interface; and changing contrast due to the compression of air. For each pair, 300 anatomical landmarks were selected by a clinical expert, with an inter-observer error of less than 1 mm [10]. We compare our method to **gsyn** [5], which is a symmetric, diffeomorphic, continuous optimisation-based registration tool, and **drop** [2], which is a discrete

---

[1] The 4D CT dataset with landmarks is freely available at `http://www.dir-lab.com`

Table 1. Results for deformable registration of inhale and exhale CT. Average target registration error (TRE) for 300 expert selected landmarks per scan pair is given in mm. Best performing algorithm per case is set in bold (for comparison: Ø TRE obtained by [11] for the same dataset is 2.13±1.8 mm). Average computation time, (maximum) degrees of freedom (d.o.f.), and harmonic energy of the deformation fields. The deformation fields of both **deeds** and **gsyn** have no negative Jacobians, the ones of **drop** exhibit a small fraction of 8.5×10^{-4}.

#	initial	**drop**	**gsyn**	**deeds**	#	initial	**drop**	**gsyn**	**deeds**
1	3.89±2.8	1.10±0.7	1.03±0.5	**0.80±0.7**	8	15.0±9.0	6.64±8.2	7.06±8.6	**2.78±3.1**
2	4.34±3.9	1.12±0.8	1.02±0.6	**0.86±0.7**	9	7.92±4.0	2.89±2.6	1.89±2.0	**1.35±0.8**
3	6.94±4.1	1.74±1.4	1.31±1.0	**1.14±0.8**	10	7.30±6.3	2.03±2.6	2.06±3.0	**1.50±1.4**
4	9.83±4.9	2.78±4.1	**1.65±1.6**	1.71±1.7	Ø	8.46±6.6	2.85±4.0	2.43±4.1	**1.60±1.7**
5	7.48±5.5	2.03±2.0	2.05±2.3	**1.77±1.7**	avg. time		8 min	29 min	20 min
6	10.9±7.0	5.20±4.5	2.50±3.3	**1.88±1.4**	d.o.f.		3.7×10^6	2.2×10^7	9.8×10^8
7	11.0±7.4	3.02±3.4	3.77±5.7	**2.21±2.3**	harm. energy		0.12	0.05	0.13

optimisation method using a B-spline deformation grid. In order to have a fair comparison, we used the same parameter settings as those chosen by the authors themselves in the EMPIRE challenge. For all approaches, 4 levels were used. As similarity metric, SAD is used for **drop** and **deeds**, and normalized cross correlation (NCC) with a radius of 2 voxels for **gsyn** (SAD is not differentiable). The dense sampling range for **deeds** is defined to be $\mathcal{L} = \{0, \pm 1, \ldots, \pm 15\}^3$ voxels, ($|\mathcal{L}| = 29791$) for the coarsest level (for datasets (1-5), with smaller deformations, the sampling range of **deeds** was decreased to $\mathcal{L}_{\max} = 10$). At each subsequent level, the range is halved. In our multi-level framework, we use cubic groups of voxels of sizes 6^3, 4^3 and 2^3. For **drop** the memory requirements (using 3.5 GB of RAM) limit us to use a sparse sampling of $\mathcal{L} = 3 \times \{0, \pm 1, \ldots, \pm 10\}$ ($|\mathcal{L}| = 61$), with a range of 24 mm. The Gaussian smoothing for gradient and deformation fields in **gsyn** is set to 3 and 1 voxels respectively. The regularisation parameters in **drop** $\lambda = 5$ and **deeds** $\alpha = 50$ were empirically chosen (intensities are in the range [0,256]); a quadratic penalty function is used for both methods.

The resulting target registration error (TRE, see Table 1) of **deeds** is 1.25 mm lower than **drop**, and 0.83 mm lower than **gsyn**. For cases with larger deformations the differences are most substantial, because the larger search space (and degrees of freedom, see Table 1) of **deeds** helps to avoids local minima. The improvements are significant, based on a Wilcoxon rank test ($p < 3 \times 10^{-4}$). The registration outcome for Case 9 is illustrated in Fig. 2, where an improved alignment using our method can clearly be seen. Another advantage of the non-iterative dense displacement sampling is the preservation of naturally occurring discontinuous motion fields. This sliding motion occurs when the lung lobes slip along their surfaces at the boundary to the rib cage and between fissures. Figure 2 gives an example, where this sliding motion is clearly preserved using **deeds** while the registration using **drop** or **gsyn** smoothes over the motion boundary.
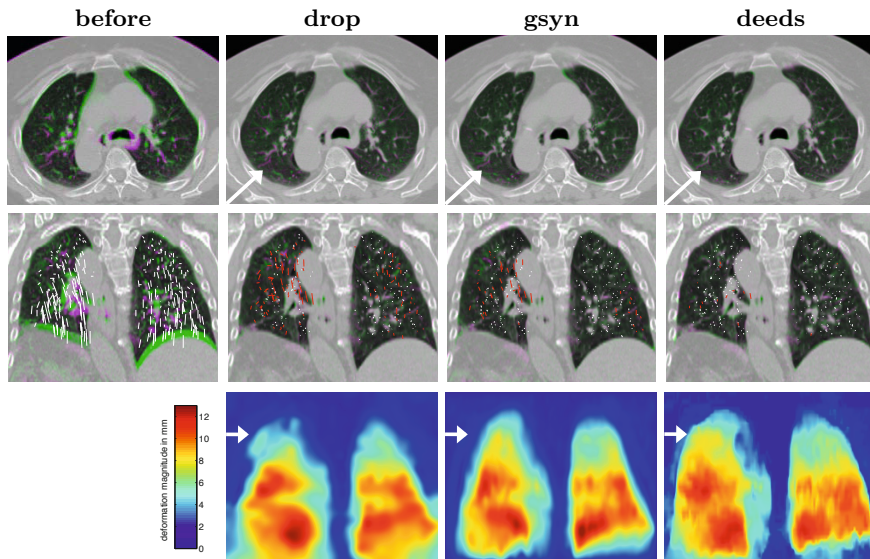
Fig. 2. Registration result for Case 9 of 4D CT dataset. The target volume is displayed in magenta, the moving volume in green (complementary colour). Axial slices before and after registration are shown in the top row. The arrow points to an improved alignment of lung vessels using our approach. The second row shows the coronal plane along with vectors indicating the registration error (errors larger than the voxel size are marked in red). The magnitudes of deformation fields is shown in the bottom row. The sliding motion of the lungs against rib cage and the heart (descending aorta) is better preserved using **deeds**. The deformation fields of **drop** and **gsyn** are too smooth close to the motion boundary at the lung surface (see arrow).

4   Conclusion and Discussion

We have introduced a novel deformable registration method **deeds** that uses discrete optimisation. A dense displacement sampling is performed for the similarity term on the highest available resolution. The regularity of the deformation field is obtained using dynamic programming on a minimum spanning tree. This ensures a globally optimal solution without the need for an iterative scheme. The algorithm is efficiently implemented in a symmetric multi-level framework yielding comparable computation time to state-of-the-art algorithms, but employing many more degrees of freedom. An average TRE of 1.60 mm was found for a challenging dataset of inhale-exhale CT scans. This is a significant improvement over the most popular discrete optimisation framework **drop** [2] (TRE=2.85 mm) and **gsyn** [5] (TRE=2.43 mm), which performed best on the recent lung registration challenge (EMPIRE). The TRE was found to be higher than the results from the EMPIRE challenge [1] for several reasons (larger deformations, lower resolution, and only one annotation per landmark). Most importantly, in contrast to the EMPIRE study, no lung segmentations were used to guide

the registration (and mask out intensities outside the lungs). Lung masks can substantially increase the registration accuracy within the lungs. However, this introduces an additional step and does not recover the full physical deformation. Our registration errors (using only inhale and exhale volumes) are smaller than the voxel size for all cases and comparable to the best performing technique on this dataset (4DLTM), which utilises all frames of the breathing cycle. In the presence of large deformations, our proposed dense displacement sampling yields a higher robustness against misregistration than **drop** and **gsyn**. Furthermore, it intrinsically deals very well with the sliding motion at the pleural interface, and hence avoids numerically more complex modelling as done e.g. in [11]. In the future, we plan to apply this new optimisation method on multi-modal datasets. Here we belief the improvements will be even more significant, due to the higher ambiguity in the similarity term.

References

1. Murphy, K., van Ginneken, B., Reinhardt, J., Kabus, S., Ding, K., Deng, X., Pluim, J.: Evaluation of registration methods on thoracic CT: The EMPIRE10 challenge. IEEE Trans. Med. Imag. 30(11), 1901–1920 (2011)
2. Glocker, B., Komodakis, N., Tziritas, G., Navab, N., Paragios, N.: Dense image registrations through MRFs and efficient linear programming. Med. Imag. Anal. 12(6), 731–741 (2008)
3. Veksler, O.: Stereo correspondence by dynamic programming on a tree. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 384–390. IEEE Computer Society (2005)
4. Felzenszwalb, P., Huttenlocher, D.: Efficient belief propagation for early vision. Int. J. Comp. Vis. 70(1), 41–54 (2006)
5. Avants, B., Epstein, C., Grossman, M., Gee, J.: Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. Med. Imag. Anal. 12(1), 26–41 (2008)
6. So, R., Tang, T., Chung, A.: Non-rigid image registration of brain magnetic resonance images using graph-cuts. Pattern Recogn. 44(10), 2450–2467 (2011)
7. Prim, R.: Shortest connection networks and some generalizations. Bell System Technical Journal 36, 1389–1401 (1957)
8. Chen, M., Lu, W., Chen, Q., Ruchala, K., Olivera, G.: A simple fixed-point approach to invert a deformation field. Medical Physics 35, 81 (2007)
9. Arsigny, V., Commowick, O., Pennec, X., Ayache, N.: A Log-Euclidean Framework for Statistics on Diffeomorphisms. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4190, pp. 924–931. Springer, Heidelberg (2006)
10. Castillo, R., Castillo, E., Guerra, R., Johnson, V., McPhail, T., Garg, A., Guerrero, T.: A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets. Phy. Med. Biol. 54(7), 1849 (2009)
11. Schmidt-Richberg, A., Werner, R., Handels, H., Ehrhardt, J.: Estimation of slipping organ motion by registration with direction-dependent regularization. Med. Imag. Anal. 16(1), 150–159 (2012)

Unbiased Groupwise Registration of White Matter Tractography

Lauren J. O'Donnell[1,2,3], William M. Wells III[3], Alexandra J. Golby[2,3], and Carl-Fredrik Westin[1,3]

[1] Laboratory for Mathematics in Imaging
[2] Golby Lab, a Surgical Brain Mapping Laboratory, and
[3] Surgical Planning Laboratory,
BWH, Harvard Medical School, Boston MA USA
`odonnell@bwh.harvard.edu`

Abstract. We present what we believe to be the first investigation into unbiased multi-subject registration of whole brain diffusion tractography of the white matter. To our knowledge, this is also the first entropy-based objective function applied to fiber tract registration. To define the probability of fiber trajectories for the computation of entropy, we take advantage of a pairwise fiber distance used as the basis for a Gaussian-like kernel. By employing several values of the kernel's scale parameter, the method is inherently multi-scale. Results of experiments using synthetic and real datasets demonstrate the potential of the method for simultaneous joint registration of tractography.

Keywords: registration, white matter, tractography, diffusion MRI.

1   Introduction

Automated medical or neuroscientific analyses of white matter tractography data, such as segmentation or labeling, creation of atlases, and measurement of tract statistics, all require initial alignment or normalization of tractography via some method. This alignment is most often performed by applying the transformations resulting from an image-based fractional anisotropy or diffusion tensor registration [18,4]. However if the eventual goal is modeling and analysis of white matter tracts, it may be advantageous to register the tracts themselves, as the quantity being optimized during registration will be closely related to the final goal. In this work we explore the possibility of driving an unbiased multi-subject registration using the trajectory data produced by streamline tractography.

In contrast to the proposed approach, to our knowledge all other methods for simultaneous joint registration of tractography have been based on alignment of pre-defined fiber bundles. These methods have required a pre-existing tractography segmentation for each subject and have thus been limited to particular structures: corticospinal tract, forceps major, cingulum and inferior longitudinal fasciculus [1]; structures resulting from an initial clustering plus expert labels [19]; left uncinate and front-occipital fasciculi [17]; and the arcuate fasciculus,

corticospinal tract, and middle cerebellar peduncles [3]. So far, methods that have performed registration using unlabeled fiber tracts from the whole brain, e.g. [9,21,12], have been limited to subject-to-subject (pairwise) registration.

In addition to tractography-based registration, our current work builds on two other categories of related work: fiber tract comparison, and groupwise image registration. Work in fiber tract clustering has led to many different metrics [16,15,10], generally based on distances computed between points along the tracts, often with conversion to fiber affinities using Gaussian kernels as in our proposed objective function. Tracts have also been analyzed via many styles of point-wise matching, for example [2,12,10,1]. In the image registration field, several groups have proposed multiple-subject unbiased and template-based image registration methods. These include entropy-based congealing methods for 2D [8] and 3D [20] that find a population central tendency image by minimizing entropy, as well as methods that estimate a population template image that is the minimum distance (in the space of diffeomorphisms) from all input images [6,4].

2   Methods

2.1   Objective Function

Our basic approach is to represent a brain or atlas by a probability distribution on trajectories. A "brain" distribution is constructed as a kernel density estimate from the tractography, and an "atlas" distribution is constructed as a mixture of the constituent brain distributions. We then choose the alignment parameters on the collection of brains by maximizing the "sharpness" of the atlas, or minimizing its entropy.

Given a distance metric D between fibers we define the probability of a fiber f, given another fiber f_j, as

$$p(f|f_j) = \frac{1}{Z} e^{-\frac{D^2(f,f_j)}{\sigma^2}} \tag{1}$$

where the distance is used as the basis for a Gaussian-like kernel with standard deviation σ, and the normalization constant Z will be discussed later. Our current choice of D is discussed below, however this can in principle work for any of the many existing fiber distances from the literature. Next, the probability of a fiber f, given the set of all fibers A and their transformations T ("the atlas"), is defined as

$$p(f|A,T) = \frac{1}{|A|} \sum_j p(f|f_j \in A, T) \tag{2}$$

where all fibers f_j in A contribute to the total probability.

The Shannon entropy H of the distribution of fibers is the expected value of the negative log-probability of the fibers. In this case the set of current transformations T has been applied to the fibers (including the transformations T_i and

T_j currently applied to fibers f_i and f_j), and we replace the expected value with the sample average value (using the weak law of large numbers).

$$H(f|A,T) = E(-log(p(f|A,T))) \tag{3}$$

$$= -\frac{1}{|A|} \sum_i log \frac{1}{|A|} \sum_j p(f_i|f_j, T_i, T_j) \tag{4}$$

We minimize the entropy as our objective function, arriving at a set of transformations T.

$$T = \arg\min_T H(f|A,T) = \arg\min_T (-\frac{1}{|A|} \sum_i log \frac{1}{|A|} \sum_j \frac{1}{Z} e^{-\frac{D^2(f_i(T_i), f_j(T_j))}{\sigma^2}}) \tag{5}$$

Note that to simplify the concept and the notation above, we have not mentioned the fact that the fibers come from several brains. This is implicitly handled in that the transformation T_i applied to fiber f_i is the same transformation that is also applied to all other fibers from that brain. We assume that Z is constant for a given value of σ, and thus does not contribute to the optimization.

2.2   Fiber Representation and Distance Function

For simplicity and computational speed, we convert the input variable-length fiber trajectories to a fixed-length representation (as also proposed by [12,14]). In practice, representing each fiber by 5 points (endpoints, midpoint, and two intermediate points) was empirically found to be effective for registration.

Using this fiber parameterization, we propose a pairwise fiber distance metric D that is related to the Hausdorff distance (the maximum of the minimum distances between pairs of closest points). We calculate D as the maximum distance between pairs of corresponding points along the fibers (i.e. the first through fifth point pairs). This fiber distance computation can thus take advantage of matrix subtractions. D is a symmetric distance that is the same between fibers (f_i, f_j) and (f_j, f_i), eliminating the issue of the classic Hausdorff measure being a directed distance. Because point ordering along the fiber is not known a priori (a fiber parameterization can equivalently start from either end), D is computed with both possible orderings and the minimum is chosen.

In practice, this method works very well with relatively nearby and corresponding fibers. For more distant fibers the point correspondence and distance measure may not be informative, a known problem with all such fiber distance metrics that have been shown to capture the local structure but poorly reflect the "true relationship" of distant fibers [16]. Luckily in our case, these uninformative large distance measures are unimportant for registration. These far-away, dissimilar, or outlier fibers are distant relative to the radius of interest defined by σ and have little effect on the objective function.

2.3   Implementation

We have implemented a full affine registration framework using a coordinate descent method. The code is a Python package that uses VTK [7], scipy [5], and

numpy [5]. For optimization we use Powell's simplex-based COBYLA (Constrained Optimization by Linear Approximation) method [13] in the scipy.optimize toolkit. The affine parameters are constrained (across subjects) as in other entropy-based congealing methods [8,20] to avoid an unnecessary overall rotation or translation of all brains, and to avoid the shrink to a point solution that artifactually reduces entropy. We require that all translation, rotation, and shear components sum to 0 over all transforms applied to the data, and that all scale factors average to 1. The COBYLA package allows definition of expected initial step sizes ρ_{beg}, and determines convergence when final step sizes ρ_{end} are under a user-provided threshold, thus we have set these ρ parameters empirically using the expected magnitudes of our transform parameters.

The σ parameter of the Gaussian kernel (eq. 1) has been tuned to enable multi-scale registration. In practice, we run several iterations of optimization, alternating translation and rotation, with an initial σ of 30mm. Next, we decrease σ to 10mm, then to 5mm, and optimize while alternating translation, rotation, scale, and shear. The computation of the fiber distances has been implemented in a multiprocessor framework. The distances are computed between a random sample of fibers from each input subject (typically 200-300), and another, smaller random sample of these fibers whose size we increase during the registration process (typically beginning with 25 or more fibers). The smaller random sample is resampled (all sampling is done without replacement) each time we change the parameters being optimized. The style of randomly sampling data points at which to compute the objective function has been successfully employed in many registration strategies [11,8,20] and in fiber clustering [15]. The terms in eq. 5 that result from comparing fibers from the same brain are neglected.

2.4   Data and Processing

N=26 healthy subjects dataset: Diffusion weighted images (DWI) scans were acquired on a 3-T GE system using an echo planar imaging sequence and the following parameters: 51 gradient directions with b=900, eight baseline scans with b=0, TR 17000 ms, TE 78 ms, FOV 24 cm, 144×144 encoding steps, 1.7 mm slice thickness. Artifacts due to eddy currents and head motion were removed by affine registration of diffusion to baseline images using FSL's linear image registration tool (FLIRT). Single-tensor streamline tractography was seeded in the entire brain of each subject in voxels with anisotropy (linear measure) > 0.2.

3   Results

We performed three registration experiments: objective function probing, synthetic data validation, and multi-subject registration.

Experiment 1: Objective Function Probing. We investigated the behavior of the multi-scale objective function under simple rotation, translation, scaling, and shear. One healthy control subject was chosen, and 2000 trajectories were

randomly sampled without replacement, twice, to generate different fixed and moving brains. A range of transformations was applied to the moving brain, and the objective function was computed using all fibers from both brains (see fig. 1). Importantly, results demonstrate that the objective function is very smooth, and that decreasing σ has the desired behavior of increasing sensitivity to small transformations.
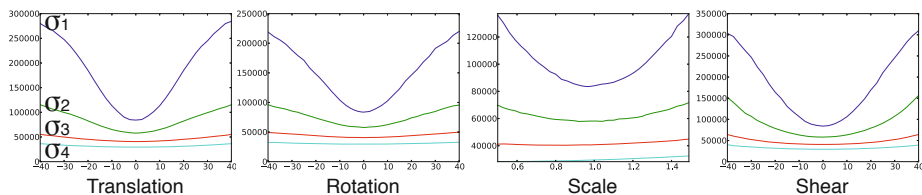
Fig. 1. Plots of the objective function under x-translation (-40 to 40mm), rotation about x (-40 to 40 degrees), scaling along x (factors of 0.5 to 1.5), and shear (-40 to 40 degrees skew about x). The curves represent different values of the multi-scale parameter σ: 5mm (top curve, blue), 10mm, 20mm, and 30mm (lowest curve, cyan).

Experiment 2: Synthetic Data Validation. Using as input one healthy control subject, we generated synthetic data as follows: 300 trajectories of length greater than 40mm were randomly sampled from the input dataset, and a randomly generated transformation was applied to these trajectories, to generate a "synthetic brain." The parameters of the random transform were: translation up to ± 20mm along each axis, scale factor from 0.85 to 1.15 in each axis, and rotation up to ± 20 degrees around each axis. To enable unambiguous computation of errors in the other parameters, shear was not included. This procedure was repeated 10 times to generate a dataset of 10 synthetic brains with known ground-truth transformations. The registration pipeline was applied to the 10 brain dataset (see fig. 2), using three levels of scale: σ of 30, 10, and 5mm; and 3 levels of randomly sampled subset fibers: 25, 50, and 75 fibers. These parameters were set empirically as a compromise between fast optimization and good convergence. Errors in the resulting parameters were measured by comparison to the ground truth applied transforms. The mean absolute errors and their standard deviations in each component were: $1.33 \pm 1.49, 1.50 \pm 1.20, 2.06 \pm 2.06$ degrees rotation; $0.62 \pm 0.456, 0.74 \pm 0.548, 2.07 \pm 0.770$ mm translation; and $0.015 \pm 0.014, 0.006 \pm 0.007, 0.017 \pm 0.015$ scale factor magnitude. The method cannot recover any mean transformation that may have been applied (e.g. if all input brains were rotated together by 30 degrees that could not be detected) so any mean transformations were removed from the ground truth transforms before calculation of the errors. The experiment ran for 48.8 minutes, spending the following amount of time at each level of scale: 6.6 minutes at 30mm, 18.1 minutes at 10mm, and 24 minutes at 5mm. (The computing environment was a 2x2.26 GHz Quad-Core Mac with 16GB of memory. Note that reported run times could be improved by coding in C rather than python, and/or increasing use of multiprocessing, however this initial implementation is a proof of concept.)
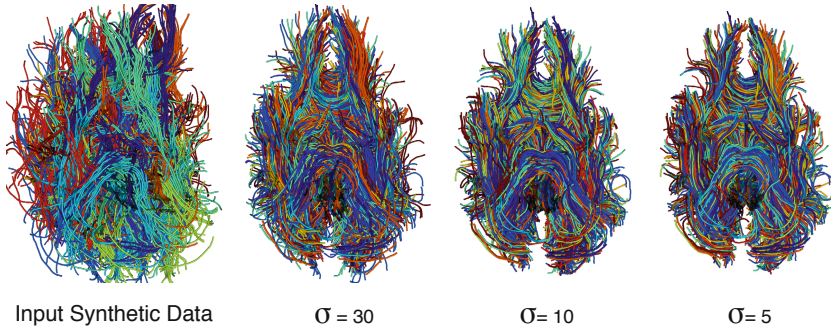
Fig. 2. Results of multi-scale registration of randomly transformed synthetic brain (n=10) dataset (inferior view). Each subject is shown in a different color. The output brains are shown after each registration scale (σ), demonstrating successful coarse-to-fine registration. The output brains appear slightly "rotated" relative to standard AC-PC orientation due to some mean component of the initial random transforms.
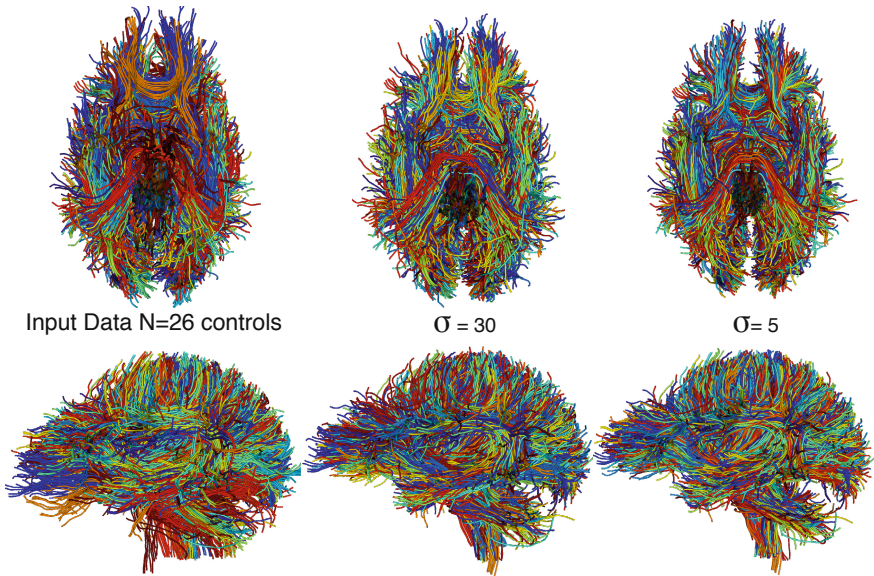
Fig. 3. Results in N=26 healthy subject dataset demonstrate successful coarse-to-fine registration (inferior and left views shown). Each subject is shown in a different color.

Experiment 3: Multi-Subject Dataset. The proposed registration method was applied to the full (N=26) healthy control multi-subject dataset (see fig. 3). The registration pipeline used 200 randomly sampled fibers of length greater than 60mm per subject, three levels of scale: σ of 30, 10, and 5mm; and 4 levels of numbers of randomly sampled subset fibers: 25, 50, 75, and 100 fibers. The method spent 48 minutes at the 30mm scale, 188 minutes at the 10mm scale, and

235 minutes at the 5mm scale. The results demonstrate successful alignment of the brains, as can be appreciated visually in fig. 3, where the output trajectories look locally similar and parallel, and the subject colors are generally mixed locally (i.e. trajectories from many subjects are neighboring).

4   Discussion and Conclusion

We have proposed a probabilistic atlas model for tractography that enables the computation of the entropy of a collection of fibers, and we have shown that registration by minimizing this entropy can successfully align the white matter in multiple subjects. Advantages of our objective function include its smoothness and the fact that any fiber outliers will have little effect. Optimization of the proposed objective, because it is based on tractography data, has the potential to enhance downstream tractography modeling and statistical analysis results. Future work will include code optimization, incorporation of higher-order deformations, and comparison of the method to other fiber- and image-based registration methods. To our knowledge, this work represents the first method for groupwise registration of whole-brain diffusion tractography data.

References

1. Caan, M., Majoie, C., van Vliet, L., van der Graaff, M., Vos, F.: Non-rigid point set matching of white matter tracts for diffusion tensor image analysis. IEEE Transactions on Biomedical Engineering 58(9), 2431–2440 (2010)
2. Corouge, I., Fletcher, P., Joshi, S., Gouttard, S., Gerig, G.: Fiber tract-oriented statistics for quantitative diffusion tensor MRI analysis. Medical Image Analysis 10(5), 786–798 (2006)
3. Durrleman, S., Fillard, P., Pennec, X., Trouvé, A., Ayache, N.: Registration, atlas estimation and variability analysis of white matter fiber bundles modeled as currents. NeuroImage 55, 1073–1090 (2011)
4. Goodlett, C., Fletcher, P., Gilmore, J., Gerig, G.: Group analysis of DTI fiber tract statistics with application to neurodevelopment. NeuroImage 45(1), 133–142 (2009)
5. Jones, E., Oliphant, T., Peterson, P., et al.: SciPy: Open source scientific tools for Python (2001), http://www.scipy.org/
6. Joshi, S., Davis, B., Jomier, M., Gerig, G.: Unbiased diffeomorphic atlas construction for computational anatomy. NeuroImage 23, 151–160 (2004)
7. Kitware: VTK: The Visualization Toolkit, http://www.vtk.org/
8. Learned-Miller, E.: Data driven image models through continuous joint alignment. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(2), 236–250 (2006)

 9. Leemans, A., Sijbers, J., De Backer, S., Vandervliet, E., Parizel, P.: Multiscale white matter fiber tract coregistration: A new feature-based approach to align diffusion tensor data. Magnetic Resonance in Medicine 55(6), 1414–1423 (2006)
10. Maddah, M., Grimson, W., Warfield, S., Wells, W.: A unified framework for clustering and quantitative analysis of white matter fiber tracts. Medical Image Analysis 12(2), 191–202 (2008)
11. Mattes, D., Haynor, D., Vesselle, H., Lewellen, T., Eubank, W.: PET-CT image registration in the chest using free-form deformations. IEEE Transactions on Medical Imaging 22(1), 120–128 (2003)
12. Mayer, A., Zimmerman-Moreno, G., Shadmi, R., Batikoff, A., Greenspan, H.: A supervised framework for the registration and segmentation of white matter fiber tracts. IEEE Transactions on Medical Imaging 30(1), 131–145 (2011)
13. Powell, M.J.D.: A direct search optimization method that models the objective and constraint functions by linear interpolation. In: Advances in Optimization and Numerical Analysis, pp. 51–67. Kluwer Academic, Dordrecht (1994)
14. O'Donnell, L., Rigolo, L., Norton, I., Westin, C.F., Golby, A.: fMRI-DTI modeling via landmark distance atlases for prediction and detection of fiber tracts. NeuroImage (2011)
15. O'Donnell, L., Westin, C.F.: Automatic tractography segmentation using a high-dimensional white matter atlas. IEEE Transactions on Medical Imaging 26(11), 1562–1575 (2007)
16. Tsai, A., Westin, C.F., Hero, A., Willsky, A.: Fiber tract clustering on manifolds with dual rooted-graphs. In: Computer Vision and Pattern Recognition, pp. 1–6. IEEE (2007)
17. Wassermann, D., Rathi, Y., Bouix, S., Kubicki, M., Kikinis, R., Shenton, M., Westin, C.-F.: White Matter Bundle Registration and Population Analysis Based on Gaussian Processes. In: Székely, G., Hahn, H.K. (eds.) IPMI 2011. LNCS, vol. 6801, pp. 320–332. Springer, Heidelberg (2011)
18. Zhang, H., Avants, B., Yushkevich, P., Woo, J., Wang, S., McCluskey, L., Elman, L., Melhem, E., Gee, J.: High-dimensional spatial normalization of diffusion tensor images improves the detection of white matter differences: an example study using amyotrophic lateral sclerosis. IEEE Transactions on Medical Imaging 26(11), 1585–1597 (2007)
19. Ziyan, U., Sabuncu, M., Grimson, W., Westin, C.F.: Consistency clustering: A robust algorithm for group-wise registration, segmentation and automatic atlas construction in diffusion MRI. Int. J. Comput. Vis. 85(3), 279–290 (2009)
20. Zöllei, L., Learned-Miller, E.G., Grimson, W.E.L., Wells, W.M.: Efficient Population Registration of 3D Data. In: Liu, Y., Jiang, T.-Z., Zhang, C. (eds.) CVBIA 2005. LNCS, vol. 3765, pp. 291–301. Springer, Heidelberg (2005)
21. Zvitia, O., Mayer, A., Shadmi, R., Miron, S., Greenspan, H.: Co-registration of white matter tractographies by adaptive-mean-shift and gaussian mixture modeling. IEEE Transactions on Medical Imaging 29(1), 132–145 (2010)

Regional Manifold Learning for Deformable Registration of Brain MR Images

Dong Hye Ye[1], Jihun Hamm[2], Dongjin Kwon[1], Christos Davatzikos[1], and Kilian M. Pohl[1]

[1] Department of Radiology, University of Pennsylvania, Philadelphia, PA, 19104
[2] Department of Computer Science, Ohio State University, Columbus, OH, 43210
Dong.Ye@uphs.upenn.edu

Abstract. We propose a method for deformable registration based on learning the manifolds of individual brain regions. Recent publications on registration of medical images advocate the use of manifold learning in order to confine the search space to anatomically plausible deformations. Existing methods construct manifolds based on a single metric over the entire image domain thus frequently miss regional brain variations. We address this issue by first learning manifolds for specific regions and then computing region-specific deformations from these manifolds. We then determine deformations for the entire image domain by learning the global manifold in such a way that it preserves the region-specific deformations. We evaluate the accuracy of our method by applying it to the LPBA40 dataset and measuring the overlap of the deformed segmentations. The result shows significant improvement in registration accuracy on cortex regions compared to other state of the art methods.

Keywords: Manifold Learning, Image Registration, Brain MRI.

1 Introduction

The analysis of deformation from non-rigid registration has become an important component in brain image applications such as morphometric analysis [1] and atlas-based segmentation [6]. To improve registration accuracy and thus the subsequent analysis, recent publications on registration [3,5,7,11] first learn the manifold capturing the neighborhood relationship of a set of images before registering individual scans. Registration then consists of determining the geodesic path between the image pairs and decomposing the deformation into a series of small deformations along that path. Since each subject moves only towards its nearby subject, the resulting deformations can be more accurate.

However, the state-of-the-art in this domain faces several challenges. First, the accuracy of the manifold in capturing the neighborhood relationship of the underlying data structure highly depends on the metric used for measuring differences between images. Current methods typically use a single metric over the entire image domain. For example, ABSORB [7] measures intensity difference of the images and GRAM [11] computes the distance based on pairwise deformations between whole brain anatomies. However, two images might be very similar in certain image regions but very different in other regions. A single

global metric generally will blur those differences and thus not accurately capture the neighborhood relations of localized brain regions. A second challenge relates to the sample size used for training the manifolds. Since manifolds are constructed directly from image samples, training based on a limited number of the image samples is likely to result in a poor approximation of the true data structure. This concern is especially relevant to medical image domain, where studies are usually limited to a few hundred samples. We address these issues by developing a deformable registration method based on regional manifolds.

Our method constructs the manifold over the entire image region in two steps. First, it separately learns the manifolds for individual brain regions and uses these manifolds to compute *region-specific* deformations. The construction of these regional manifolds is now based on metrics that are much more sensitive to local variations within that region than a single global metric. In addition, the anatomical variation within a specific region is smaller compared to one captures in the entire image domain so that our method can faithfully learn regional manifolds with a relatively small number of samples. In the second step, our method learns the manifold over the entire image domain so that moving along geodesics of that manifold does not interfere with the deformations inferred from the regional manifolds. Specifically, we use a Markov Random Field model to produce smooth deformation maps across the entire image domain while preserving the region-specific deformations. In other words, our approach determines the optimal geodesic path over the entire image region by gradually warping localized brain regions according to the regional manifolds.

We demonstrate the advantage of our method by performing atlas-based segmentation on LPBA40 dataset. The results show significant improvement in registration accuracy on cortex regions in terms of overlap score compared to other state-of-the art registration methods.

2   Regional Manifold Learning Based Registration

We now describe our Regional Manifold Learning based Registration (RMLR) as illustrated in Fig.1. RMLR first defines a set of regions of interests (ROIs) in training images. Then, RMLR independently learns the manifolds for individual regions and finds the region-specific deformations constrained by the regional manifolds. Next, RMLR learns the manifold for the whole brain image based on regional manifolds. Finally, RMLR determines the deformation in the entire image domain based on the global manifold while preserving region-specific deformations. We outline each step in more detail in the remainder of this section.

Defining the ROIs. To define the ROIs, we first set the ROIs in the template image I_T. Specifically, we separate the entire image domain Ω into R image regions $\{\Omega^r : r = 1, \ldots, R\}$ so that the union of all regions is a subset of Ω (i.e. $\cup_{r=1}^{r=R} \Omega^r \subset \Omega$) and each region does not overlap with another region, (i.e. for $\forall r, \nexists s \neq r : \{\Omega^r \cap \Omega^s \neq \emptyset\}$). We then automatically find the corresponding regions in training images $\{I_i : i = 1, \ldots, N\}$ using a non-rigid registration [10].
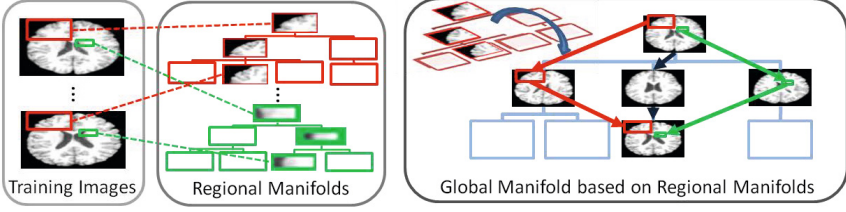
Fig. 1. Illustration of the proposed method: It first defines ROIs and then learns the manifolds for each ROI. Next, the global manifold is constructed based on these regional manifolds. Then, the localized brain regions will be warped to their nearest neighbors based on the regional manifolds (red and green), however adjacent regions will deform in mutually compatible ways maintaining the smoothness of the overall deformation based on the global manifold (blue).

Computing the Regional Manifolds. For each Ω^r, we learn the regional manifolds as in GRAM [5], which was originally proposed for learning the manifold of an entire image domain. In order to reduce the risk of boundary artifacts during registration, we compute the manifold for a larger image block $B^r \supset \Omega^r$. Let $\phi_{B^r}(I_i, I_j)$ be the diffeomorphic deformation which maps the image block in I_i to the corresponding block in I_j. Then, the distance $d_{B^r}(I_i, I_j)$ between blocks is defined as the weighted sum of intensity difference and field smoothness [3,5]:

$$d_{B^r}(I_i, I_j) := \sum_{u \in B^r} ||I_i \circ \phi_{B^r}(I_i, I_j)(u) - I_j(u)||_2^2 + \lambda_r|| \bigtriangledown \phi_{B^r}(I_i, I_j)(u)||_2^2, \quad (1)$$

where λ_r is a weighting parameter between intensity difference and field smoothness term, and $||\cdot||$ is the $L2$ norm. To reduce the computational burden in computing the distances for all pairs of blocks, we use the symmetric diffeomorphic registration [10] and compute $d_{B^r}(I_i, I_j)$ only for $i < j$. Also, we empirically set λ_r so that two terms have the same maximum value over all pairs of images.

Based on these pairwise distances, we construct k-NN graph for each region whose nodes correspond to the image blocks. Heuristically, we set the smallest value that makes the k-NN graph connected as k. From the k-NN graph, we find the shortest paths from the block in the template I_T to all other image blocks. Then, the shortest paths from one root node to the rest form a spanning tree with respect to B^r. We consider this spanning tree M_{Ω^r} as the regional manifold for Ω^r that represents the neighborhood relations in the specific region.

Registering the ROIs. After computing the regional manifold, we construct the region-specific deformation based on this regional manifold. Between two image blocks in I_i and I_j, we choose the shortest path $[i, p_1, \ldots, p_l, j]$ within the graph M_{Ω^r} and define $\hat{\phi}_{B^r}(I_i, I_j) := \phi_{B^r}(I_i, I_{p_1}) \circ \cdots \circ \phi_{B^r}(I_{p_l}, I_j)$ by concatenating the corresponding deformations between neighbor blocks along that path. $\hat{\phi}_{B^r}(I_i, I_j)$ is diffeomorphic because composition operator preserves this property. The region-specific deformation $\hat{\phi}_{\Omega^r}(I_i, I_j)$ is simply the $\hat{\phi}_{B^r}(I_i, I_j)$ restricted to the region Ω^r. This process is performed on all R regions independently, producing region-specific deformation fields.

Computing the Global Manifold. Given registration results in the ROIs, we compute the manifold for the entire image domain Ω. Similar to regional manifold computation, we measure the distance based on pairwise deformations $\phi_\Omega(I_i, I_j)$ between I_i and I_j. To preserve pre-computed region-specific deformations while maintaining the smoothness of overall deformation, it is desirable to compute the globally diffeomorphic deformation in such a way that it keeps the region-specific deformations. Toward this, we determine $\phi_\Omega(I_i, I_j)$ using an MRF based registration method [4]. A typical MRF model optimizes the following energy function:

$$E := \sum_{s \in \mathcal{V}} \theta_s(x_s) + \sum_{(s,t) \in \mathcal{E}} \theta_{st}(x_s, x_t), \qquad (2)$$

where $\mathcal{V}$ is a set of nodes on the image, $\mathcal{E}$ is a set of edges between neighbor nodes, and x_s is the label of node $s \in \mathcal{V}$. Each label x_s corresponds to a displacement vector $\mathbf{v}(x_s)$ by which s moves to a new position. The unary term $\theta_s(x_s)$ represents the data cost of assigning label x_s at node s in terms of image dissimilarity. The smoothness term $\theta_{st}(x_s, x_t)$ penalizes the cost of label discrepancy between two neighboring nodes s and t, i.e. $\theta_{st}(x_s, x_t) := \gamma_{st} \cdot \min\{||\mathbf{v}(x_s) - \mathbf{v}(x_t)||_1, T_{st}\}$, with γ_{st} being a regularization constant and T_{st} being a threshold for truncation. In order to preserve the region-specific deformations, we define a modified unary term $\hat{\theta}_s(x_s)$ as following:

$$\hat{\theta}_s(x_s) := \begin{cases} L(1 - \delta(x_s - \hat{x}_s)) & \text{for } s \in \cup_{r=1}^{r=R} \Omega^r \\ 1 - NCC(x_s) & \text{for } s \in \Omega \setminus \cup_{r=1}^{r=R} \Omega^r, \end{cases} \qquad (3)$$

where L is very large, $\hat{x}_s$ is the pre-defined label from region-specific deformations, $\delta(\cdot)$ is a delta function, and NCC is a normalized cross correlation. We optimize this energy model via tree re-weighted message passing method [8] with $\gamma_{st} = 3$ and $T_{st} = 20$. In order to guarantee the diffeomorphism, we constrain the displacement at each node not to exceed 40% of the spacing between two adjacent nodes [2]. The MRF-based registration for each pair of subjects gives us ϕ_Ω which is diffeomorphic and preserves the region-specific deformations.

Based on these MRF registrations, we construct the k-NN graph whose nodes represent the images. From this graph, we find a spanning tree from one root node to all images and consider this spanning tree M_Ω as the global manifold.

Registering the Whole Brain Images. We now find the global deformation $\hat{\phi}_\Omega(I_i, I_j)$ between I_i and I_j constrained by the global manifold M_Ω. We denote $[i, G_1, \ldots, G_m, j]$ as the shortest path between I_i and I_j in M_Ω. Then, $\hat{\phi}_\Omega(I_i, I_j)$ is defined as $\hat{\phi}_\Omega(I_i, I_j) := \phi_\Omega(I_i, I_{G_1}) \circ \cdots \circ \phi_\Omega(I_{G_m}, I_j)$, where ϕ_Ω is the MRF propagated deformation from the previous step. We now show that $\hat{\phi}_\Omega(I_i, I_j)$ preserves the region-specific deformation $\hat{\phi}_{\Omega^r}(I_i, I_j)$ in Ω^r.

We prove the claim by contradiction. First, in Ω^r, the deformation ϕ_Ω between neighbor nodes on M_Ω is defined by $\hat{\phi}_{\Omega^r}$ (according to M_{Ω^r}) because our MRF registration preserves the region-specific deformation. Then, the deformation for any path on M_{Ω^r} with the same start and end node $[b, \ldots, b]$ is the identity as

M_{Ω^r} is loop-free and $\hat{\phi}_{\Omega^r}$ is symmetric. Thus, the deformation of path passing through a node twice, such as $[a, \ldots, b, \ldots, b, \ldots, c]$, is equivalent to the deformation defined by the path $[a, \ldots, b, \ldots, c]$ without that loop. Now, suppose that there is an image pair with $\hat{\phi}_{\Omega^r}(I_i, I_j) \neq \hat{\phi}_{\Omega}(I_i, I_j)$ on Ω^r. Let $[i, p_1, \ldots, p_l, j]$ and $[i, G_1, \ldots, G_m, j]$ be the shortest paths in M_{Ω^r} and M_{Ω}, respectively. Then, we define $[i, G'_1, \ldots, G'_n, j]$ as the corresponding equivalent path of $[i, G_1, \ldots, G_m, j]$ in M_{Ω^r} by first replacing two neighboring nodes $[G_q, G_{q+1}]$ with the corresponding path in M_{Ω^r} and then removing all loops according to the previous observation. We note that a node in either path $[i, p_1, \ldots, p_l, j]$ or $[i, G'_1, \ldots, G'_n, j]$ can only appear once in that path. If now $\hat{\phi}_{\Omega^r}(I_i, I_j) \neq \hat{\phi}_{\Omega}(I_i, I_j)$ on Ω^r, then two paths have to differ in at least one position k, i.e. $p_k \neq G'_k$. This implies that the loop-free graph structure M_{Ω^r} has a loop as there are two unique loop-free paths between node i and j. As this contradicts our assumption of M_{Ω^r}, it follows that $\hat{\phi}_{\Omega}(I_i, I_j)$ always has to be equivalent to $\hat{\phi}_{\Omega^r}(I_i, I_j)$ on Ω^r.

3   Experiments on LPBA40 Dataset

We measure the accuracy of our RMLR method by applying it to the LPBA40 dataset [9]. The dataset consists of 40 linearly aligned brain images each with 54 manually labeled segmentations. From those 40 scans, we empirically choose one scan as a template. We then determine the registration accuracy by aligning all other subjects to the template and measuring the overlap between the aligned and template segmentations. For comparison, we measure the accuracy of direct diffeomorphic Demons [10] which aligns the images without any manifold learning. In addition, we compare our RMLR method with whole brain GRAM [5] which learns the manifold based on a single metric over the entire image domain. For all three methods, we used the same registration parameters with three levels of resolution and the smoothing kernel size of 1.5. The remainder of this section discusses the experimental results in further detail.

Choosing Regions. The proposed method is independent of the choices of ROIs. In this paper, we try out the two schemes shown in Fig.2. First, we define regions simply as cubes over the entire image domain as shown in Fig.2(a). The

(a) cubic regions   (b) structure regions
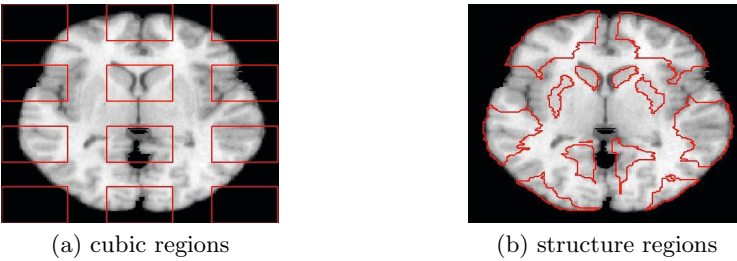
Fig. 2. The mid-axial slice of the template image with (a) cubic regions and (b) structure regions. Red boundaries indicate ROIs.

(a) region-specific deformation   (b) whole brain deformation
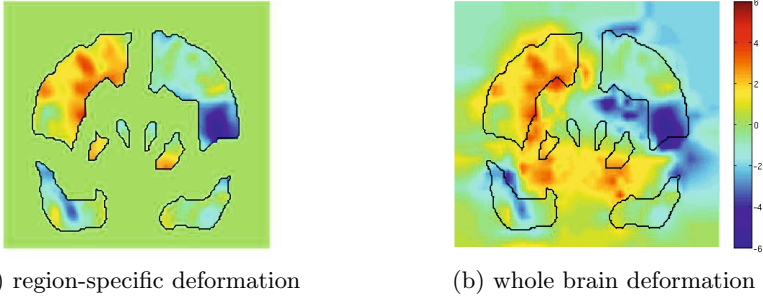
Fig. 3. y-displacement in a coronal view for (a) region-specific and (b) whole brain deformations from RMLR. Black boundaries represent the ROIs. Notice that whole brain deformation is globally smooth while preserving region-specific deformations.

$151 \times 188 \times 136$ image is divided into 36 cubic regions whose sizes are $37 \times 32 \times 32$. In order to separate cubic regions, we empirically set a gap of 20-voxels between regions. The second scheme uses the segmentation of template image I_T to define regions according to anatomical brain structures. As illustrated in Fig.2(b), we specify 10 regions including frontal lobe, parietal lobe, occipital lobe, temporal lobe, and sub-cortical regions of left and right hemispheres. Each structure region is refined through erosion to guarantee the non-overlap between regions.

Comparing Regional and Global Deformations. First, we visually compare the region-specific deformation with the whole brain deformation from our RMLR method. Fig.3(a) shows an example of region-specific deformations based on regional manifolds. We only display y-displacement vector field in a mid-coronal slice. Black boundaries represent structure ROIs including frontal lobe, temporal lobe and subcortical regions. Initially, there is only zero-displacement vector field outside the ROIs. Fig.3(b) illustrates the whole brain deformation based on the global manifold which is constructed from regional manifolds. Note that the whole brain deformation is globally smooth and preserves the region-specific deformation from the Fig.3(a). This supports our claim that MRF registration produces the globally smooth deformation and the region-specific deformation can be preserved during the MRF registration and composition along the path on the global manifold.

Comparing the RMLR and GRAM. Next, we compare the regional manifolds from RMLR with the manifold produced by GRAM based on a single metric over a whole brain image. To do so, we visualize the shortest paths for the left hippocampus and right angular gyrus in Fig.4(a) and Fig.4(b), respectively. Upper row represents the GRAM manifold and lower row represents the regional manifolds from RMLR. First, we note that the paths are the same for both regions according to the GRAM manifold (upper) but not so for RMLR (lower). This indicates that the anatomical variation for one region is generally different from that of another region and the manifold based on a single global metric may not capture this local variation with limited samples. Furthermore, regional manifolds reflect more gradual changes in both regions in terms of hippocampus

(a) left hippocampus   (b) right angular gyrus
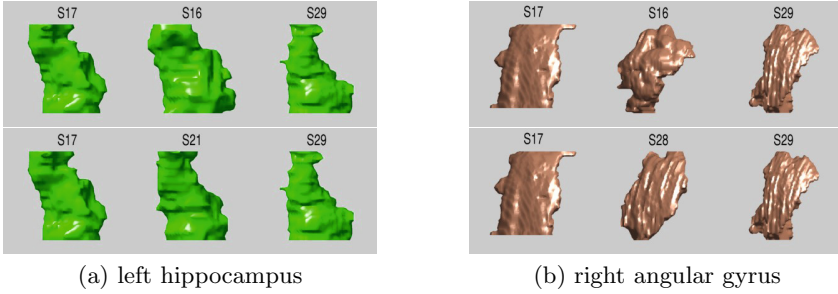
Fig. 4. Shortest path for (a) left hippocampus and (b) right angular gyrus. The upper row is the path by GRAM and the lower row is the path by regional manifolds. The numbers on top is the subject ID. Notice gradual decreases in hippocampus size and sulcal depth in the paths of regional manifolds compared with those from GRAM.

size and gyrus appearance compared to the path from GRAM manifold. This also confirms that regional manifold can better capture the anatomical variation in the specific region with a small number of samples.

Measuring DICE Scores. To measure registration accuracy, we compute the DICE score between manual segmentation and atlas-based segmentation from all registration methods. The average DICE score over all 54 labels on the original dataset is 57.4% and the score after direct diffeomorphic Demons is 72%. Whole brain GRAM slightly improves the score to 73.1%. RMLR increases the score to 75.1% with the cubic and to 75.2% with the structural region setting. This shows that our RMLR achieves improvement in average DICE score over all labels and that this improvement is somewhat independent towards the region selection schemes. For further comparison, we display in Fig 5 the 25, 50 and 75 percentiles of the DICE scores for the four different registration methods with respect to selected structures. RMLR produces scores with higher median and lower variation than GRAM or Demons. In particular, RMLR achieves statistically significant improvement in cortex regions ($p < 0.05$). If we use the scores of Demons as an indicator regarding the degree of difficulty in registering the
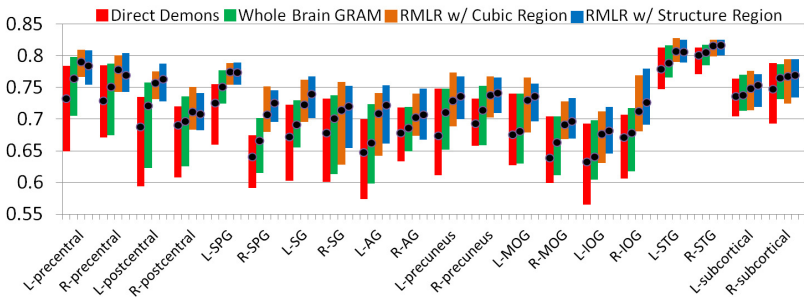
Fig. 5. Structure-specific DICE scores. Upper and lower bar represent 75 and 25 percentiles of DICE scores, respectively. Midpoint indicates the median.

regions then the regional manifolds provides the biggest improvements in difficult regions, such as angular gyrus, and only slightly impacted the results in easy regions, such as the temporal gyrus and subcortical regions. In summary, these results agree with our initial intuition that manifolds based on a single metric over the entire image domain are not as accurate in capturing local brain variations than ones that are based on regional manifolds.

4   Conclusion

In this paper, we proposed a deformable registration based on learning the manifolds of individual brain regions. Our method first learns the manifolds for specific regions and then computes region-specific deformations from these manifolds. We then determine deformations for the entire image domain by learning the global manifold while preserving the region-specific deformations via a MRF model. Experimental results on the LPBA40 dataset show that the proposed method can significantly improve registration accuracy compared to direct pairwise or whole brain manifold learning based registration methods. In a future work, we will investigate an adaptive way of our method for overlapping ROIs.

References

1. Ashburner, J., Friston, K.J.: Voxel-based morphometry–the methods. Neuroimage 11(6 Pt 1), 805–821 (2000)
2. Choi, Y., Lee, S.: Injectivity conditions of 2D and 3D uniform cubic B-spline functions. Graphical Models 62(6), 411–427 (2000)
3. Gerber, S., Tasdizen, T., Fletcher, P.T., Joshi, S.C., Whitaker, R.T.: Manifold modeling for brain population analysis. Med. Imag. Anal. 14(5), 643–653 (2010)
4. Glocker, B., Komodakis, N., Tziritas, G., Navab, N., Paragios, N.: Dense image registration through MRFs and efficient linear programming. Med. Imag. Anal. 12(6), 731–741 (2008)
5. Hamm, J., Ye, D.H., Verma, R., Davatzikos, C.: GRAM: A framework for geodesic registration on anatomical manifolds. Med. Imag. Anal. 14(5), 633–642 (2010)
6. Heckemann, R.A., Hajnal, J.V., Aljabar, P., Rueckert, D., Hammers, A.: Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. Neuroimage 33(1), 115–126 (2006)
7. Jia, H., Wu, G., Wang, Q., Shen, D.: Absorb: Atlas building by self-organized registration and bundling. Neuroimage 51(3), 1057–1070 (2010)
8. Kolmogorov, V.: Convergent tree-reweighted message passing for energy minimization. Tech. Rep. MSR-TR-2004-90, Microsoft Research (MSR) (September 2004)
9. Shattuck, D.W., Mirza, M., Adisetiyo, V., Hojatkashani, C., Salamon, G., Narr, K.L., Poldrack, R.A., Bilder, R.M., Toga, A.W.: Construction of a 3D probabilistic atlas of human cortical structures. Neuroimage 39(3), 1064–1080 (2008)
10. Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Symmetric Log-Domain Diffeomorphic Registration: A Demons-Based Approach. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 754–761. Springer, Heidelberg (2008)
11. Wolz, R., Aljabar, P., Hajnal, J.V., Hammers, A., Rueckert, D.: ADNI: Leap: learning embeddings for atlas propagation. Neuroimage 49(2), 1316–1325 (2010)

Estimation and Reduction of Target Registration Error

Ryan D. Datteri and Benoît M. Dawant

Department of Electrical Engineering and Computer Science,
Vanderbilt University, Nashville, TN 37235, USA
{ryan.d.datteri,benoit.dawant}@vanderbilt.edu

Abstract. Fiducial-based registration is often utilized in image guided surgery because of its simplicity and speed. The assessment of target registration error when using this technique, however, is difficult. Although the distribution of the target registration error can be estimated given the fiducial configuration and an estimation of the fiducial localization error, the target registration error for a specific registration is uncorrelated with the fiducial registration error. Fiducial registration error is thus an unreliable predictor of the target registration error for a particular case. In this work, we present a new method to estimate the quality of a fiducial-based registration and show that our measure is correlated to the target registration error and that it can be used to reduce registration error caused by fiducial localization error. This has direct implication on the attainable accuracy of fiducial-based registration methods.

Keywords: Image registration, registration circuits, rigid registration, fiducial registration, image guided surgery, registration error.

1   Introduction

Fiducial-based registration is an important technique in Image Guided Surgery (IGS). It is often utilized to align image information to the surgical space in an operating room. In this context, fiducial markers are attached to the patient and an image is acquired. The physical location of the fiducial markers in the operating room is obtained as well as the location of the markers in the image and the two point sets (fiducial configurations) are registered to each other. Error in identifying the correct location of the individual fiducials, called Fiducial Localization Error (FLE) [1] may occur which causes error in the registration between the image and surgical space. An analytical solution for the distribution of errors in fiducial registration has been proposed [1, 2, 3], but this solution does not permit the assessment of the target registration error in a particular case. Fiducial Registration Error (FRE) is often used as a surrogate for the Target Registration Error (TRE) that is the quantity of clinical interest [4]. Unfortunately, it has been shown [5] that TRE and FRE are uncorrelated. FRE is thus an unreliable predictor of registration accuracy.

In this work, we propose a method to estimate the quality of a registration that correlates with TRE and therefore produces a value that correlates with the true registration accuracy at a target location. The next section describes our technique, which we call AQUIRC for Assessing Quality Using Image Registration Circuits. In

the results section, simulation results we have produced to demonstrate the correlation between our measure and TRE are presented and future work and applications are discussed.

2   General Algorithm

This algorithm was first proposed in [6] for global atlas selection and was utilized to estimate the quality of intensity-based rigid image registration in [7]. Briefly, AQUIRC builds on the idea of registration circuits which was proposed as a consistency measure by Woods et al. [8] and Holden et al. [9]. Here, a registration circuit involves three fiducial configurations A, B, and C and three transformations T_{AB}, T_{BC}, and T_{CA}. The configurations' coordinates differ by the assumed localization error of the individual fiducials. As discussed by Fitzpatrick [10], using only one registration circuit can lead to an underestimation of registration error because the error made along one edge in the circuit may correct error introduced from a separate edge in the circuit.

In this work, we expand upon the idea of a registration circuit to multiple circuits. We start with a set of fiducial configurations and compute pair-wise registrations between all elements in the set, creating a complete graph as shown in Figure 1. The complete graph of registrations is similar to what is done by Christensen [11]. In [11], however, the set consisted of medical images and the
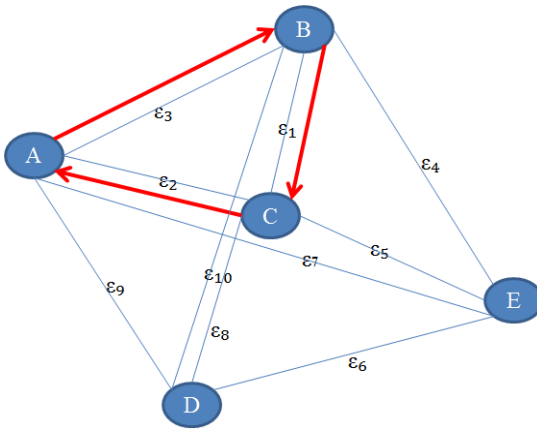
Fig. 1. Example complete graph with one circuit shown in red arrows

complete graph was used as an overall measure of quality for a registration algorithm, rather than as a method to determine the quality of individual registrations as we have done here. If our initial set contains N fiducial configurations (i.e., the same set of fiducial markers but with the position perturbed by the assumed fiducial localization error) the graph contains $\binom{N}{2}$ edges. With each edge in this graph, we associate an initially unknown measure of registration quality called ε that we wish to solve for. There are $\binom{N}{3}$ unique registration circuits that can be formed from a complete graph (we have used registration circuits of size 3; the circuit size can be increased to form more registration circuits but this was not explored here).

Next, we define a measure of registration error that can be computed across a circuit. Here, to compute this error, we select a target point or set of points in A, say X. We then compute the transformed point(s) X′ as $X' = T_{AB}(T_{BC}(T_{CA}(X)))$. We note the important fact that the order in which transformations are composed is critical and that this order differs from the originally proposed registration circuit in [8, 9]. The quality of

registrations across circuit A, B, C, is then defined as $E_C = dissimilarity(X, X')$. The value, E_C, is affected by the error of three registrations, i.e., the registration error between A and B, the registration error between B and C, and the registration error between C and A. With only one circuit the contribution of each component cannot be computed. It can, however, be estimated with more than one circuit. To achieve this we make the assumption that each registration affects the quality measure multiplicatively, i.e., $\varepsilon_{ABC} = \varepsilon_A * \varepsilon_B * \varepsilon_C$ or, $\log(\varepsilon_{ABC}) = \log(\varepsilon_A) + \log(\varepsilon_B) + \log(\varepsilon_C)$. An additive model may also be applicable but was not tested in this work. Computing this expression for all possible circuits and rearranging them in matrix form, we obtain

$$
\begin{bmatrix}
1\,1\,1\,0\ldots0 \\
1\,0\,1\,1\ldots0 \\
1\,1\,0\,1\ldots0 \\
0\,1\,1\,1\ldots0 \\
. \\
. \\
.
\end{bmatrix}
\begin{bmatrix}
\log(\varepsilon_1) \\
\log(\varepsilon_2) \\
\log(\varepsilon_3) \\
. \\
. \\
. \\
\log(\varepsilon_{\binom{N}{2}})
\end{bmatrix}
=
\begin{bmatrix}
\log(E_{c_1}) \\
\log(E_{c_2}) \\
\log(E_{c_3}) \\
. \\
. \\
. \\
\log(E_{c_{\binom{N}{3}}})
\end{bmatrix}
\tag{1}
$$

in which E_{C_i} is defined as the $dissimilarity(X, X')$ value around circuit i. This expression can be rewritten as $\overline{\overline{P}} \log(\overline{\varepsilon}) = \log(\overline{E_c})$. As a result of the multiplicative assumption, $\log(\overline{\varepsilon})$ can be solved for using a linear least squares solution

$$
\log(\overline{\varepsilon}) = \left(\overline{\overline{P}}^T \overline{\overline{P}}\right)^{-1} \overline{\overline{P}}^T \log(\overline{E_c})
\tag{2}
$$

and finally solving for $\overline{\varepsilon}$

$$
\overline{\varepsilon} = e^{\log(\overline{\varepsilon})}
\tag{3}
$$

We are currently working on a proof of conditions on the registration circuits for when $\overline{\overline{P}}$ is full rank and therefore $\left(\overline{\overline{P}}^T \overline{\overline{P}}\right)$ is invertible. Experimentally $\overline{\overline{P}}$ has been observed to be full rank when $N \geq 5$. We define $\overline{\overline{P}}$ to be all unique circuits in the graph of size 3.

There are multiple ways to define the circuits that are utilized to create the $\overline{\overline{P}}$ matrix. In this work, we utilize the set of unique circuits of size 3 in the graph. For example, for the three nodes A, B, and C in Figure 1, we consider only one circuit e.g., $T_{AB}(T_{BC}(T_{CA}))$.

2.1   AQUIRC Applied to Fiducial-Based Registration

For this work, we utilize the fiducial registration method that minimizes the FRE between two sets of points. This fiducial registration method is standard and uses singular-value decomposition as proposed by Arun et al [12]. We also define the function $dissimilarity(X, X')$ to be the TRE(X, X') where TRE(X, X') is defined as the Euclidean distance between X and X' because, as explained below, our set X

contains a single element. In this particular study, we define two spaces: the *image space* and the *surgical space*, to mimic the situation of a typical point-based registration problem where pre-operative images need to be registered to the patient in the operating room. We define X as the target point and we select that point in image space. As discussed below, we introduce a known transformation between image and surgical space that can be large and we perturb the position of the fiducials in image space.

3   Experiments and Results

To test our algorithm we repeat an experiment that was performed by Fitzpatrick which showed analytically and experimentally that FRE and TRE are uncorrelated [5]. Using the same experiment we show that our quality measure is correlated to the TRE.

3.1   Experiments

In Experiment 1 in [5], Fitzpatrick simulates an actual Deep Brain Stimulation case with 4 fiducials and a target location in the deep brain. The location of the four fiducials x_1, x_2, x_3, and x_4 as well as the target position x_t are:

$$x_1 = \begin{bmatrix} 197 \\ 217 \\ 115 \end{bmatrix}, \quad x_2 = \begin{bmatrix} 109 \\ 225 \\ 121 \end{bmatrix}, \quad x_3 = \begin{bmatrix} 83 \\ 139 \\ 127 \end{bmatrix}, \quad x_4 = \begin{bmatrix} 202 \\ 132 \\ 130 \end{bmatrix}, \quad x_t = \begin{bmatrix} 144 \\ 155 \\ 57 \end{bmatrix}$$

Following the same simulation model as in [5], we first apply a rotation R and translation t to the location of the fiducials x_1, x_2, x_3, and x_4 as well as to the target position x_t, which results in the corresponding positions y_1, y_2, y_3, y_4, and y_t. We consider the fiducials x_1, x_2, x_3, and x_4 and x_t to be in the image space while the rotated and translated fiducials y_1, y_2, y_3, y_4, and y_t are considered to be in the surgical space. Again, as in [5], we set the rotation R to be 10, 20, and -30 degrees about the x, y, and z axes and we set the translation t to be (7, -10, 100) mm (which was chosen as an arbitrary mis-registration in [5]).

We then perturb the location of the fiducials in image space using a fiducial localization error drawn from a random distribution with zero mean and variance of FLE/3, where FLE is set to 1mm. This is done N-1 times to create a set of N-1 perturbed fiducial configurations (as seen in the left of Figure 2). We then compute all pair-wise fiducial registrations between each of the N-1 fiducial configurations in the image space as well as between these and the unperturbed rotated fiducial configuration in the surgical space, creating the complete graph of registrations necessary to run our algorithm which we use to calculate the ε value for each of the registrations. There are three quantities of importance in this work: TRE, FRE, and ε. These three values are calculated for the registrations between the fiducial configurations in the image space and the fiducial configuration in the surgical space, which results in N-1 values for TRE, FRE, and ε (this is represented as the red links in Figure 2). FRE is defined as the root mean square distance between the fiducial points

in image space and the points transformed from image space to surgical space using the computed transformations. For each run, we thus produce N-1 FRE values. Similarly, TRE is defined as the Euclidean distance between the target position in surgical space and the position of the target transformed from image space to surgical space. We repeat this process in a Monte Carlo simulation; creating thousands of image space fiducial configurations with randomized FLEs.

In the first experiment we utilize an N of size 30 over 1000 simulations. We calculate the correlation between the TRE values and ε values and the FRE and ε values. In the second experiment we test the effect N has on our results by varying the size of N from 5 to 25 using 5000 simulations for each value of N. Again, we calculate the correlation between TRE and the ε values for each value of N. Finally, in the third experiment, we test the ability of our algorithm to improve the TRE in a fiducial-based registration scenario. To do this we use an N of 30 over 1000 simulations. For each simulation we consider the registration between the 29 image space fiducials and the surgical space fiducial and for all 29 registrations calculate the mean TRE, the min TRE, the max TRE, the TRE of the fiducial configuration that AQUIRC identifies as being of the highest quality, and the TRE of the fiducial configuration with the minimum FRE.
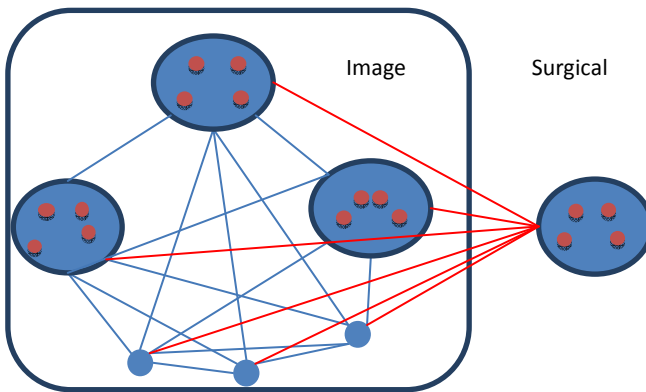
Fig. 2. Diagram of the experiment methodology. The image space contains N-1 fiducial configurations that are created by adding FLE to the original fiducial locations. The surgical space contains the rotated and translated fiducials. Each set of fiducial configurations are registered to every other configuration, both in image space and in surgical space. The red links represent the registrations for which we calculate the TRE, FRE and ε values.

3.2   Results

The results of experiment 1 are shown in Figures 3. In the left of Figure 3 we show a scatter plot of the FRE and TRE values (the points in both scatter plots were reduced to a random sampling of 1000 data points for better visualization). As can be seen, we produce a correlation that is very similar to the one found in [5], with an r = -0.0012, which is not statistically significant (p = 0.8351). In the right of Figure 3 we show the correlation between our algorithm's ε value and TRE. In this case there is a statistically significant correlation of r = 0.6086 and a p < 0.001. The results of

experiment 2 are shown in Figure 4. By increasing the number of fiducial configurations that we utilize in the image space we can increase the correlation between TRE and ε, producing better estimations of the quality of error in fiducial registrations. The results of experiment 3 are shown in Figure 5; it shows the practical utility of our algorithm. If, for every simulation, we choose the configuration with the lowest ε value, the TRE is reduced by a statistically significant 0.1347 mm when compared to the mean TRE value and is reduced by a statistically significant 0.1367 when compared to the TRE of the fiducial configuration with the minimum FRE. The figure also shows the mean and standard deviation of the TRE values when the configurations with the max and min TRE values are selected at every run.
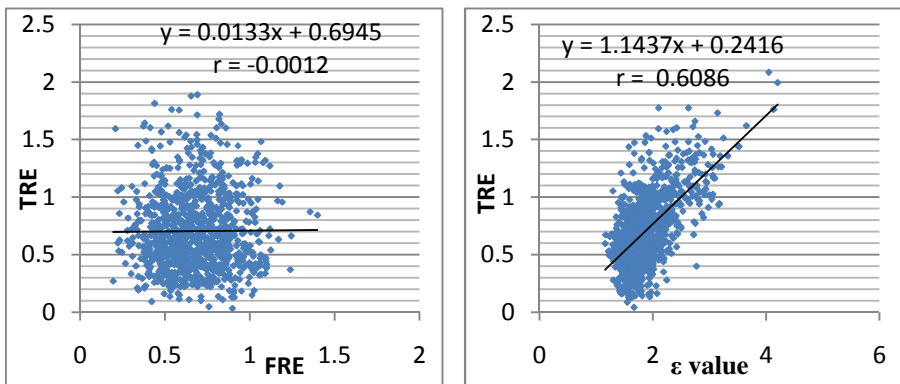
Fig. 3. Left: Scatter plot of the TRE and FRE values. Right: Scatter plot of the TRE and ε values.
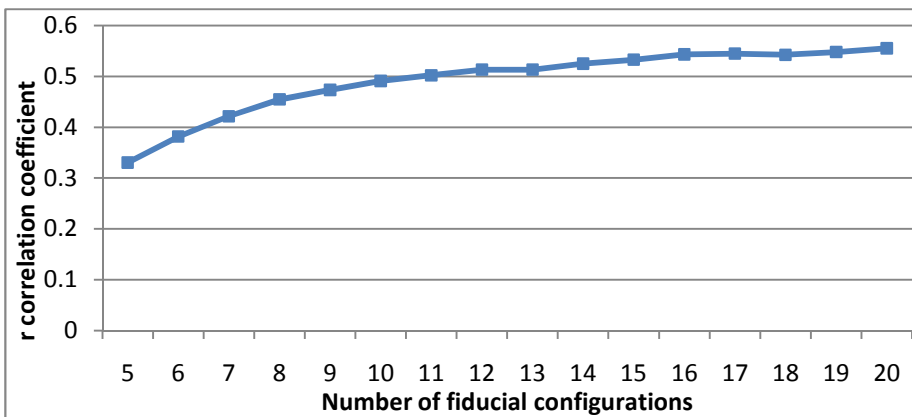
Fig. 4. Correlation between TRE and ε as a function of the number of fiducial configurations utilized in the surgical space.
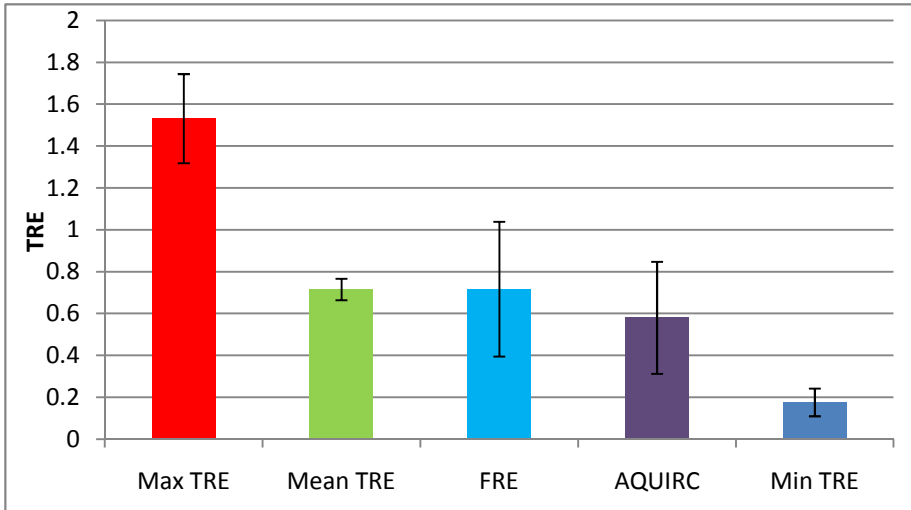
Fig. 5. Bar graph of the mean value and standard deviation of the TRE value across 1000 simulations. For each simulation we calculate the mean TRE of all 30 fiducial configurations, the min TRE of all 30 fiducial configurations, the max TRE of all 30 fiducial configurations, the TRE of the fiducial configuration that AQUIRC identifies as being of the highest quality, and TRE of the fiducial configuration with the minimum FRE.

4   Discussions/Future Work

In this work we have introduced a method that produces a measure of registration quality that is correlated to the target registration error. We are unaware of any other published work that describes a technique that is able to do so. We have shown that the number of fiducial configurations that are utilized in the complete graph of registrations affects the quality of our algorithm's estimation and the correlation between our measure and TRE increases as the number of configuration increases. Most importantly for practical applications, we also show that by choosing the configuration that our algorithm identifies as producing the best registration, we can reduce the average TRE.

To use our algorithm in practice, all that is needed is to acquire the location of the fiducial markers in the image space as well as in surgical space. The location of the markers in image space can then be randomly perturbed by an FLE that is representative of the error that naturally occurs when attempting to identify the coordinates of the markers (alternatively, the position of the markers in surgical space could be perturbed). These sets of markers can then be registered together to form a complete graph of registrations and AQUIRC is applied. The ε value of the unperturbed configuration, i.e., the position of the fiducials selected by the end user, can then be compared to the distribution of ε values produced by our algorithm. If the unperturbed ε value is large compared to the perturbed ε values, the end user could be warned of a potential registration problem.

In the future, we will investigate various types of models that could better represent the combination of error that occurs when combining multiple transformations than the multiplicative model we have used in this work. We will also explore further the distribution of ε values and attempt to define statistical tests that would permit to quantify the quality of a particular registration.

As discussed earlier, the order in which transformations are composed when computing the registration error across a circuit is important. In fact, if X' is computed as $X' = T_{AB}(T_{BC}(T_{CA}(X)))$, we have not observed a correlation between the TRE and ε values. The theoretical reasons for this observation have not yet been elucidated and are under investigation. If successful, this algorithm would provide end users with quantitative measures of accuracy for a particular registration. This would represent a major advance in the field of fiducial-based registration.

References

1. Fitzpatrick, J.M., West, J.B., Maurer Jr., C.R.: Predicting error in rigid-body point-based registration. IEEE Transactions on Medical Imaging 17, 694–702 (1998)
2. Fitzpatrick, J.M., West, J.B.: The distribution of target registration error in rigid-body point-based registration. IEEE Transactions on Medical Imaging 20, 917–927 (2001)
3. Danilchenko, A., Fitzpatrick, J.M.: General approach to first-order error prediction in rigid point registration. IEEE Trans. Med. Imaging 30, 679–693 (2010)
4. Suess, O., Picht, T., Kuehn, B., Mularski, S., Brock, M., Kombos, T.: Neurosurgery without rigid pin fixation of the head in left frontotemporal tumor surgery with intraoperative speech mapping. Operative Neurosurgery 60(suppl. 2), 330–338 (2007)
5. Fitzpatrick, J.M.: Fiducial registration error target registration error are uncorrelated. In: SPIE. Medical Imaging: Visualization, Image-Guided Procedures, and Modeling, vol. 7261(1), pp. 1–12 (2009)
6. Datteri, R.D., Asman, A.J., Landman, B.A., Dawant, B.M.: Estimation of Registration Quality Applied to Multi-Atlas Segmentation. In: MICCAI (2011)
7. Datteri, R.D., Dawant, B.M.: Estimation of Rigid-Body Registration Quality Using Registration Networks. In: Proc. SPIE, Medical Imaging 2011: Image Processing (2011)
8. Woods, R.P., Grafton, S.T., Holmes, C.J., Cherry, S.R., Mazziotta, J.C.: Automated image registration: I. General methods and intrasubject, intramodality validation. J. Comput. Assist. Tomogr. 22, 139–152 (1998)
9. Holden, M., Hill, D.L.G., Denton, E.R.E., Jarosz, J.M., Cox, T.C.S., Rohlfing, T., Goodey, J., Hawkes, D.J.: Voxel similarity measures for 3D serial MR brain image registration. IEEE Transactions on Medical Imaging 19, 94–102 (2000)
10. Fitzpatrick, J.M.: Detecting failure, assessing success. In: Hajnal, J.V., Hill, D.L.G., Hawkes, D.J. (eds.) Medical Image Registration, pp. 117–139. CRC Press, Florida (2001)
11. Christensen, G.E., Johnson, H.J.: Invertibility and transitivity analysis for nonrigid image registration. J. Electron. Imaging 12, 106–117 (2003)
12. Arun, K.S., Huang, T.S., Blostein, S.D.: Least-Squares Fitting of two 3-D Point Sets. IEEE Trans. Pattern Anal. 9(5), 698–700 (1987)

A Hierarchical Scheme for Geodesic Anatomical Labeling of Airway Trees

Aasa Feragen[1], Jens Petersen[1], Megan Owen[2], Pechin Lo[1,3],
Laura H. Thomsen[4], Mathilde M.W. Wille[4],
Asger Dirksen[4], and Marleen de Bruijne[1,5]

[1] Department of Computer Science, University of Copenhagen, Denmark
[2] David R. Cheriton School of Computer Science, University of Waterloo, Canada
[3] Department of Radiology, University of California, Los Angeles, USA
[4] Lungemedicinsk afdeling, Gentofte Hospital, Denmark
[5] Biomedical Imaging Group Rotterdam, Erasmus MC, The Netherlands
{aasa,phup,marleen}@diku.dk, megan.owen@uwaterloo.ca

Abstract. We present a fast and robust supervised algorithm for labeling anatomical airway trees, based on geodesic distances in a geometric tree-space. Possible branch label configurations for a given tree are evaluated based on distances to a training set of labeled trees. In tree-space, the tree topology and geometry change continuously, giving a natural way to automatically handle anatomical differences and noise. The algorithm is made efficient using a hierarchical approach, in which labels are assigned from the top down. We only use features of the airway centerline tree, which are relatively unaffected by pathology.

A thorough leave-one-patient-out evaluation of the algorithm is made on 40 segmented airway trees from 20 subjects labeled by 2 medical experts. We evaluate accuracy, reproducibility and robustness in patients with Chronic Obstructive Pulmonary Disease (COPD). Performance is statistically similar to the inter- and intra-expert agreement, and we found no significant correlation between COPD stage and labeling accuracy.

Keywords: airway tree labeling, geodesic matching.

1 Introduction

Computed Tomography (CT) is an important tool in the analysis of diseases affecting the airways. Using image segmentation methods, three-dimensional models of the airway surfaces can be constructed, and their dimensions measured. Such measurements are, however, dependent on the location in which they are made [4], so solving the problem of finding anatomically corresponding positions in different airway trees is crucial to robustly compare measurements across patients. It can be solved by assigning anatomical names to the airway tree branches. This is nontrivial, since the topology of the anatomical airway

Fig. 1. *Left:* The method takes only the airway centerline tree as input. *Middle-right:* Airway trees are frequently topologically different, while geometric differences are small.

tree changes from person to person[1], and the segmented trees have differences introduced by noise, including missing and spurious branches.

Several types of airway branch labeling algorithms have appeared previously. Van Ginneken et al. [11] and Lo et al. [5] use machine learning on branch features, with additional assumptions on airway tree topology. Among the features used are branch length, radius, orientation, cross-sectional shape and bifurcation angle. Branch length and radius are sensitive to the presence of structural noise or diseases like cystic fibrosis, tuberculosis or Chronic Obstructive Pulmonary Disease (COPD), and assumptions on tree topology make these methods vulnerable to topological differences. Tschirren et al. use association graphs [10] for pairs of airway trees, which incorporate information from both trees, such that maximal cliques in the association graph induce branch matchings between the original graphs. This is a combinatorial construction, although it can depend on the geometric properties of the initial trees. Such a separation of geometric and combinatorial properties can be problematic, as airway trees are often geometrically and visually similar despite being combinatorially different, as in Fig. 1. Feragen et al. [2] label airways based on geodesics (shortest paths) in a space of trees. Their tree-space is highly non-linear and has no known efficient algorithm for geodesic computation, making labeling of a complete airway tree infeasible.

We present a novel supervised method for automatic airway branch labeling, based on shortest paths in a space of geometric trees [1]. This tree-space is less general than that of Feragen et al. [2], but allows geodesics to be computed in polynomial time [6]. Possible label configurations on an unlabeled tree are evaluated based on distances to all trees in a training set of labeled trees. In tree-space, both topology and geometric branch attributes are allowed to change continuously, and we can thus compare trees with different topology and branch geometry. The only feature used is the airway centerline tree, see Fig. 1. The method does not depend directly on the branches identified by the segmentation, but rather on a subtree spanning the labeled branches, defined in Fig. 2, where branches may be concatenations of branches from the originally segmented tree. As a consequence, the method is less sensitive to structural noise such as false or missing branches than methods that work only on the originally segmented

---

[1] Tree topology describes how tree branches are connected. The tree topologies in our data are plotted in http://image.diku.dk/aasa/miccai_supplemental.tar.gz

Fig. 2. *Left:* Given a configuration of leaf labels on an airway centerline tree, extract the subtree spanning the leaf labels and prune off the rest, giving *the subtree spanning the labels*, a geometric leaf-labeled tree which can be compared to the training data. *Right:* After assigning a set of labels, each label is moved to the branch in the segmented airway tree closest to the root which is not part of the subtree spanning the other labels.

branches. The method is implemented in a hierarchical fashion, making it sufficiently fast to be of practical use. A thorough leave-one-patient-out evaluation of the algorithm is made on a set of 40 segmented airway trees from 20 subjects different stages of COPD (healthy-severe), labeled by 2 medical experts. We evaluate accuracy, reproducibility and robustness to disease stages.

2   Anatomical Branch Labeling

Airway branch labels correspond to the division of the lung into compartments: LMB and RMB lead to the left and right lungs; LUL, RUL, L4+5, R4+5, LLB, RLL lead to the lobes; and R1-R10, L1-L10 lead to (up to) 10 segments in each lung. In addition, intermediate branch names appear in the literature, whose presence depends on the topology of the anatomical airway tree. However, if the segment branch labels and the airway tree structure are known, the remaining branch labels can be reconstructed trivially. For this reason, a *leaf-labeled airway tree* (where the leaf labels are segment labels) is equivalent to a *labeled airway tree*. Thus we focus on evaluating the assignment of segment labels in this paper.

Methodology. Each airway tree was normalized by person height as an isotropic scaling parameter. Each branch was represented by 6 landmark points sampled equidistantly along the centerline, translated so that the first landmark point was placed at the origin. Thus, each branch e is represented by a vector $x_e \in \mathbb{R}^{15}$.

For an arbitrary unlabeled airway tree T, we attach a set of 20 leaf labels corresponding to the 20 segmental bronchi, named $X = \{L1, ..., L10, R1, ..., R10\}$. Our training set consists of n airway trees which have been labeled by two experts in pulmonary medicine. We extract the *subtree spanning the labels*, as defined in Fig. 2, and obtain $2n$ leaf-labeled trees $\mathfrak{T} = \{T_1^1, \ldots, T_n^1, T_1^2, \ldots, T_n^2\}$.

Given an unlabeled airway tree T, we proceed as follows. Denote the set of branches in T by E. A labeling of T is a map $L: X \to E$. We only consider labelings where the leaf labels are all attached to *leaves* in the subtree spanning the labels, i.e., we do not consider labelings where two leaf labels are attached to branches on the same path to the root. Given such a labeling L, extract the subtree T_L of T spanning the labels. For each labeled tree $\tilde{T}$ in our training set and each T_L, we compute the shortest-path distance $d(\tilde{T}, T_L)$ between the trees
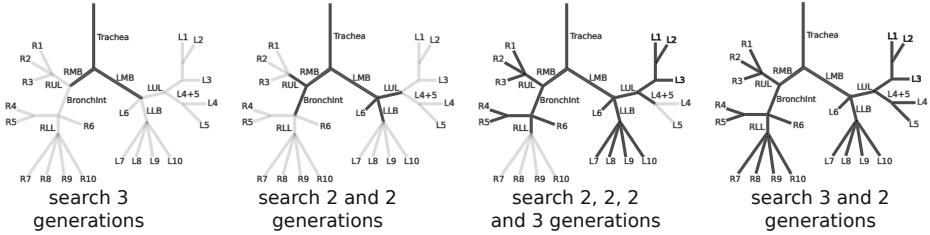
Fig. 3. Hierarchical labeling: In each step, search through 2 or 3 generations of branches (as indicated in the figure) to find an optimal alignment of a set of labels, obtaining a leaf-labeled subtree of the segmented airway tree similar to the trees shown in black. The real tree topology may differ; the figure only illustrates the stepwise hierarchy.

$\tilde{T}$ and T_L in the tree-space defined below. We find a labeling of T by choosing the labeled tree T_{labeled} among the T_L that satisfies:

$$T_{\text{labeled}} = \arg\min_{T_L} \sum_{\tilde{T} \in \mathfrak{T}} d(\tilde{T}, T_L). \tag{1}$$

Ideally, we would search through the whole airway tree T, test all admissible configurations T_L of the 20 segment leaf labels and select the one that optimizes (1). However, for an airway tree with 100 branches, the search space size is on the order of 100^{20}, which is too large to handle. In order to ensure computational feasibility, we choose a hierarchical subtree approach, where labels of different generation are added subsequently, see Fig. 3. In the first step, 2 generations below the trachea are searched for the optimal configurations of the RMB, LMB labels. In the second step, 2 and 2 generations below the RMB and LMB, resp., are searched for the optimal configurations of {RUL, BronchInt, L6, LLB, LUL}. In the third step, 2, 2, 2 and 3 generations below RUL, BronchInt, LLB, and LUL are searched for the optimal configurations of {R1-R5, L1-L3, L4+5 L7-L10}. In the final step, 3 and 2 generations below RLL, R4+5 are searched for optimal configurations of {R4, R5, R7-R10}.

In particular, we treat more shallow branches as leaves in the first steps of the algorithm, and work our way down to the segments. In each step of the hierarchical label placement, we pick the optimal branches for the given set of labels and backtrace each label through the path to the root, see Fig. 2.

Tree-Space. The tree-space used in this paper is a straight-forward generalization of the tree-space from [1], generalizing single-dimensional shape vectors on the branches to multi-dimensional ones. Any two trees are joined by a shortest path through this space, whose length defines a distance (a metric) on tree-space.

Each point in tree-space is a leaf-labeled tree, with the leaves labeled by some fixed set X. Each edge in a leaf-labeled tree can be represented by a partition of X into the leaves descending from the edge, and the remaining leaves (including the root), see Fig. 4. Then a tree will uniquely correspond, as follows, to a vector in $(\mathbb{R}^{15})^S$, where S is the number of possible partitions of X. Each consecutive set

Correlation of labeling success with COPD stage

root

e1

e2

R1 R2 R3

e1 = {R1,root},{R2,R3}
e2 = {R1,root,R2},{R3}
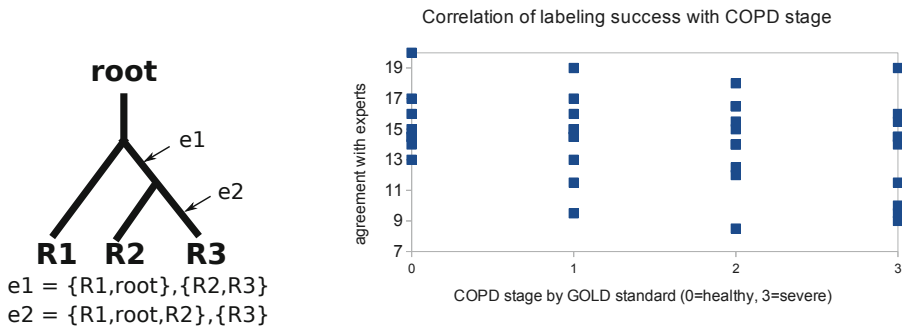
Fig. 4. *Left:* Tree edges are defined by partitions on the leaf label set. *Right:* The number of correctly assigned branch labels per segmented airway versus COPD stage.

of 15 coordinates corresponds to a possible partition of X. If the edge associated with that partition appears in the tree, then those 15 coordinates will be its branch vector, and otherwise they are all 0. Certain edges can never appear in a tree together (e.g., an edge that splits $\{R1, R2\}$ off from the rest of the tree and an edge that splits $\{R1, R3\}$ off), so not all vectors are possible. *Tree-space* is precisely those vectors in $(\mathbb{R}^k)^S$ that correspond to trees. Thus, tree-space is a subset of Euclidean space. The shortest-path distance between two trees is the shortest path between them that remains fully within this restricted subspace, with the length of the path being measured in the ambient Euclidean space using the Euclidean metric. An analytic formula for this distance does not exist, but it can be computed recursively in polynomial time. See [6] for details and code.[2]

3   Experimental Results

Data. We work with a set of 40 airway tree centerlines obtained from low-dose (120 kV and 40 mAs) pulmonary CT scans from the Danish lung cancer screening trial [7]. The images came from 20 subjects scanned at two different times, with an average interval of 5 years. There were 5 asymptomatic subjects and 5 from each of 3 different patient groups with mild, moderate and severe COPD according to the GOLD standard [9]. The images were segmented, centerlines extracted and branching points detected all automatically as described in [8].

The 40 airway trees were manually labeled by two experts in pulmonary medicine, who assigned segment labels $L1$ - $L10$ and $R1$ - $R10$ to the airway trees; the remaining labels in Fig. 3 were deduced from these. This was done using in-house developed software, simultaneously showing the segmented airway and centerline, which can be rotated, panned and zoomed, as well as a CT cross-section perpendicular to and centered on any given point of the airway.

---

[2] Code:
http://vm1.cas.unc.edu/stat-or/webspace/miscellaneous/provan/treespace/

Table 1. Labeling results. *Agreement with two experts* measures % of agreement of the automatic labeling with the experts' labeling in the cases where the experts agree. *Average intra-expert agreement* and *Automatic reproducibility* measure the reproducibility of the experts' and automatic labelings, respectively, on pairs of scans of the same patient. The automatic method does not always assign all labels; this happens when the search subtree does not have sufficiently many leaves. Plots of airway trees with attached labels as well as tables with the complete branch labeling can be found at http://image.diku.dk/aasa/miccai_supplemental.tar.gz.

Label	% Agreement with expert 1	% Agreement with expert 2	% Avg. agreement with expert	% Inter-expert agreement	% Avg. intra-expert agreement	% Automatic reproducibility	% Agreement with two experts	# Found by automatic labeling
R1	84.62	69.23	76.92	80.00	92.50	80.00	81.25	39
R2	82.05	74.36	78.21	75.00	95.00	85.00	90.00	39
R3	84.62	79.49	82.05	85.00	95.00	85.00	85.29	39
R4	92.50	80.00	86.25	87.50	82.50	90.00	91.43	40
R5	92.50	82.50	87.50	87.50	82.50	90.00	94.29	40
R6	100.00	92.50	96.25	97.50	85.00	95.00	94.87	40
R7	60.53	92.11	76.32	60.00	82.50	85.00	91.67	38
R8	39.47	84.21	61.84	45.00	60.00	65.00	77.78	38
R9	52.63	68.42	60.53	52.50	57.50	60.00	76.19	38
R10	50.00	60.53	55.26	52.50	55.00	65.00	66.67	38
L1	85.00	72.50	78.75	67.50	57.50	75.00	96.30	40
L2	85.00	75.00	80.00	75.00	57.50	75.00	93.33	40
L3	82.50	80.00	81.25	70.00	70.00	70.00	96.43	40
L4	65.00	65.00	65.00	95.00	92.50	55.00	68.42	40
L5	65.00	65.00	65.00	95.00	95.00	60.00	68.42	40
L6	100.00	100.00	100.00	100.00	100.00	95.00	100.00	40
L7	50.00	65.00	57.50	42.50	47.50	80.00	76.47	40
L8	50.00	62.50	56.25	47.50	50.00	80.00	73.68	40
L9	55.00	50.00	52.50	50.00	50.00	65.00	70.00	40
L10	55.00	60.00	57.50	55.00	45.00	65.00	77.27	40
Trachea	100.00	100.00	100.00	100.00	100.00	100.00	100.00	40
LMB	100.00	100.00	100.00	100.00	100.00	100.00	100.00	40
LUL	100.00	100.00	100.00	100.00	100.00	100.00	100.00	40
LB4+5	97.50	95.00	96.25	97.50	97.50	95.00	97.44	40
LLB	100.00	100.00	100.00	100.00	100.00	100.00	100.00	40
RMB	100.00	100.00	100.00	100.00	100.00	100.00	100.00	40
RUL	90.00	90.00	90.00	90.00	85.00	90.00	100.00	40
BronchInt	100.00	100.00	100.00	100.00	100.00	100.00	100.00	40
RLL	92.50	95.00	93.75	97.50	97.50	90.00	94.87	40
Avg segment	71.57	73.92	72.74	71.00	72.63	76.00	83.49	39.5
Avg total	79.70	81.32	80.51	79.48	80.43	82.59	88.35	39.6

Labeling Results. The labeling was implemented in MATLAB, using tree distance computations implemented in Java. For airway trees with 150 branches on average, the whole labeling takes, roughly, 5 minutes per tree running on a single 2.40 GHz processor on a laptop with 8 GB RAM.

The labeling was tested in a leave-one-patient-out fashion. Thus for each airway, the training set was made up of 38 airway trees from other patients, with each tree labeled separately by the two medical experts, giving a total of 76 training airway trees. The results of the labeling are shown in table 1, along with a comparison of the two expert labelings.

In order to test reproducibility of the expert and automatic labels, the two CT scans of each subject were registered using the image registration method described in [3], and the labeled airway trees were compared in a common coordinate system. Expert 1, Expert 2 and the automatic algorithm reproduced 14.0, 15.1 and 15.2 labels per subject on average. The automatic algorithm was not significantly different from the average expert ($p = 0.51$ in a paired t-test).

On average, the automatic labeling agreement with an expert is 72.8% on the segment branches, which is not significantly different from the average inter-expert agreement of 71.0% ($p = 0.75$ in a paired t-test). Fig. 4 shows labeling performance stratified by COPD stage. Spearman's correlation test shows no significant correlation between the average agreement with an expert and the severity of COPD ($\rho = -0.22$, $p = 0.18$).

4   Discussion and Conclusion

Higher percentages are reported in the literature: 97.1%, 90%, 83% on all branch labels in [10, high dose CT], [11], [5]; 77% on segment labels [5]. These methods use fewer than 20 segment labels and/or more intermediate (easier) labels, and reject uncertain labels using a threshold. On average, only 71%, 93% and 83% of the given label set [10], [11], [5] and at most 77% of the given segment label set [5] are assigned, whereas we almost always assign all 20 segment labels. The 97.1% success rate [10] is among branches that have been labeled identically by three experts. Taking unassigned labels into account in [5,11,10], these do not appear to perform better, and do not test reproducibility. We, on the other hand, perform just as well and reproducibly as a medical expert on our dataset.

Our evaluation, which is thorough compared to previous work, gives detailed insight into the difficulties of the labeling problem. It is noteworthy that the experts and the automatic method perform well in different parts of the airway tree. In particular, the automatic method is far more reproducible in parts of the airway tree where the experts have difficulties, e.g., the lower left lobe (L7-L10).

The fact that the method as presented always assigns all segment labels if possible, makes it sensitive to missing branches and increases our false positive rate on difficult branches. This could be tackled by introducing label probabilities based on the geodesic airway tree distances, and thus assigning fewer labels.

The hierarchical scheme of Fig. 3 may cause difficulties with rare topologies. This could be handled by a more refined hierarchical labeling scheme, particularly one informed by an analysis of where the experts performed better. The labeling is sensitive to mistakes made above the segment level. This could be improved by label probabilities; however, the algorithm rarely makes such mistakes.

We present a new supervised method for anatomical branch labeling of airway trees, based on geodesic distances between airway trees in tree-space. Using the distances, the algorithm evaluates how well a suggested branch labeling fits with a training set of labeled airway trees, and chooses the optimal labeling. The labeling performance is robust in patients with COPD, and is comparable to that of two experts in pulmonary medicine. As it only uses branch centerlines and tree topology, we expect it to generalize to other datasets. Its reproducibility and robustness in patients with COPD makes it highly suitable for clinical use.

References

1. Billera, L.J., Holmes, S.P., Vogtmann, K.: Geometry of the space of phylogenetic trees. Adv. in Appl. Math. 27(4), 733–767 (2001)
2. Feragen, A., Lo, P., Gorbunova, V., Nielsen, M., Dirksen, A., Reinhardt, J.M., Lauze, F., de Bruijne, M.: An airway tree-shape model for geodesic airway branch labeling. In: MICCAI WS on Math. Found. Comp. Anat. (2011)
3. Gorbunova, V., Sporring, J., Lo, P., Loeve, M., Tiddens, H., Nielsen, M., Dirksen, A., de Bruijne, M.: Mass preserving image registration for lung CT. MedIA (2012)
4. Hasegawa, M., Nasuhara, Y., Onodera, Y., Makita, H., Nagai, K., Fuke, S., Ito, Y., Betsuyaku, T., Nishimura, M.: Airflow Limitation and Airway Dimensions in Chronic Obstructive Pulmonary Disease. Am. J. Respir. Crit. Care Med. 173(12), 1309–1315 (2006)
5. Lo, P., van Rikxoort, E.M., Goldin, J., Abtin, F., de Bruijne, M., Brown, M.: A bottom-up approach for labeling of human airway trees. In: MICCAI Int. WS. Pulm. Im. Anal. (2011)
6. Owen, M., Provan, J.S.: A fast algorithm for computing geodesic distances in tree space. ACM/IEEE Trans. Comp. Biol. Bioinf. 8, 2–13 (2011)
7. Pedersen, J., Ashraf, H., Dirksen, A., Bach, K., Hansen, H., Toennesen, P., Thorsen, H., Brodersen, J., Skov, B., Døssing, M., Mortensen, J., Richter, K., Clementsen, P., Seersholm, N.: The Danish randomized lung cancer CT screening trial. Overall design and results of the prevalence round. J. Thor. Onc. 4(5), 608–614 (2009)
8. Petersen, J., Nielsen, M., Lo, P., Saghir, Z., Dirksen, A., de Bruijne, M.: Optimal Graph Based Segmentation Using Flow Lines with Application to Airway Wall Segmentation. In: Székely, G., Hahn, H.K. (eds.) IPMI 2011. LNCS, vol. 6801, pp. 49–60. Springer, Heidelberg (2011)

9. Rabe, K.F., Hurd, S., Anzueto, A., Barnes, P.J., Buist, S.A., Calverley, P., Fukuchi, Y., Jenkins, C., Rodriguez-Roisin, R., van Weel, C., Zielinski, J.: Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: Gold executive summary. Am. J. Respir. Crit. Care Med. 176(6), 532–555 (2007)
10. Tschirren, J., McLennan, G., Palágyi, K., Hoffman, E.A., Sonka, M.: Matching and anatomical labeling of human airway tree. TMI 24(12), 1540–1547 (2005)
11. 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)

Initialising Groupwise Non-rigid Registration Using Multiple Parts+Geometry Models

Pei Zhang[1], Pew-Thian Yap[1], Dinggang Shen[1], and Timothy F. Cootes[2]

[1] Department of Radiology and Biomedical Research Imaging Center (BRIC)
The University of North Carolina at Chapel Hill, USA
peizhang@email.unc.edu, {ptyap,dgshen}@med.unc.edu
[2] Imaging Sciences, School of Cancer and Enabling Sciences,
The University of Manchester, UK
timothy.f.cootes@manchester.ac.uk

Abstract. Groupwise non-rigid registration is an important technique in medical image analysis. Recent studies show that its accuracy can be greatly improved by explicitly providing good initialisation. This is achieved by seeking a sparse correspondence using a parts+geometry model. In this paper we show that a single parts+geometry model is unlikely to establish consistent sparse correspondence for complex objects, and that better initialisation can be achieved using a set of models. We describe how to combine the strengths of multiple models, and demonstrate that the method gives state-of-the-art performance on three datasets, with the most significant improvement on the most challenging.

1 Introduction

Finding consistent correspondences across sets of images is a challenging problem, with applications in many areas, particularly in constructing statistical models of shape or appearance [1,2]. A promising solution is groupwise non-rigid image registration [3,4,5], where the correspondence is usually determined by minimising some objective function. A common choice of initialisation for groupwise registration is affine transformation. However, our recent studies have shown that it is insufficient when registering images of large local shape variations and repeating structures [6,7]. Hence, more sophisticated approaches to initialisation have been developed [8,9,6,7].

Although implementation varies, these algorithms share the same idea, that is, deliberately finding a sparse set of corresponding points and using them to initialise groupwise registration. In [9], it was shown that a small set of manually selected parts can be used to build a parts+geometry model [10,11,12,13] capable of giving a good sparse correspondence. Later in [6,7], such human intervention was shown to be unnecessary—the set of "good" parts can be automatically obtained. Instead of using parts+geometry models, Langs et al. [8] explored using a Point Distribution Model [1] in a weakly supervised fashion. The model is iteratively estimated so as to minimise the description length of the feature points across the images, leading to the optimal correspondence.

A problem with the above methods is that they all attempt to use a *single* model to achieve the good sparse correspondence. Due to the limited amount of sample images and the imperfection of the algorithm, it is unlikely for a single model to capture every

(a) Model A (b) Consistent matches   (c) Failures

(d) Model B (e) Failures   (f) Consistent matches
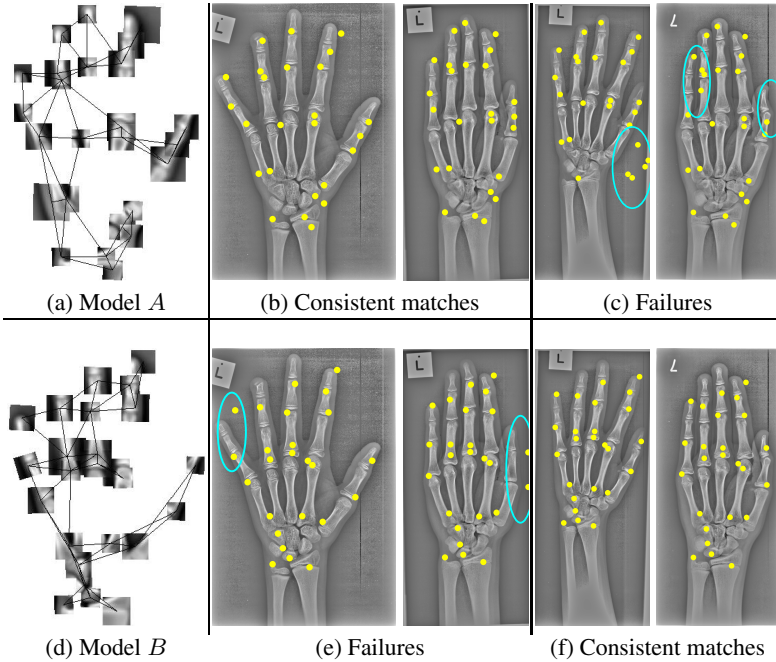
Fig. 1. Left column: two parts+geometry models. Right four columns: matches of the models. Top row: model A finds consistent matches on two images but fails on the other two. Bottom row: model B gives consistent matches on those where model A fails, but fails where model A works well. Failures are indicated by cyan ellipses.

possible variation of the object, particularly those of complex structures. As a result, the learned model may fail on some of the images. See Fig. 1 for an example.

Figure 1 also suggests that it is always possible to learn a model that can deal well with a subset of images. Different models that work well with different subsets are likely to complement each other. By combining the best result from each model, we may achieve better initialisation than just using a single model.

In this paper we explore using multiple parts+geometry models to initialise groupwise non-rigid registration—a *multi-model* initialisation scheme. We use our previous approach [6] to generate a population of reasonably good models. We then apply the algorithm described below to these models to obtain the desired correspondence.

2   Initialisation with Multiple Parts+Geometry Models

Let the term *pattern* refer to the sparse set of points found on an image by a particular model, which thus defines the correspondence for that image to a reference frame for that model. When using multiple models we obtain multiple patterns on each image. As shown in Fig. 1, some patterns are apparently better than the others. By replacing those poor patterns with good ones, we can modify and improve the correspondence defined by the single model.
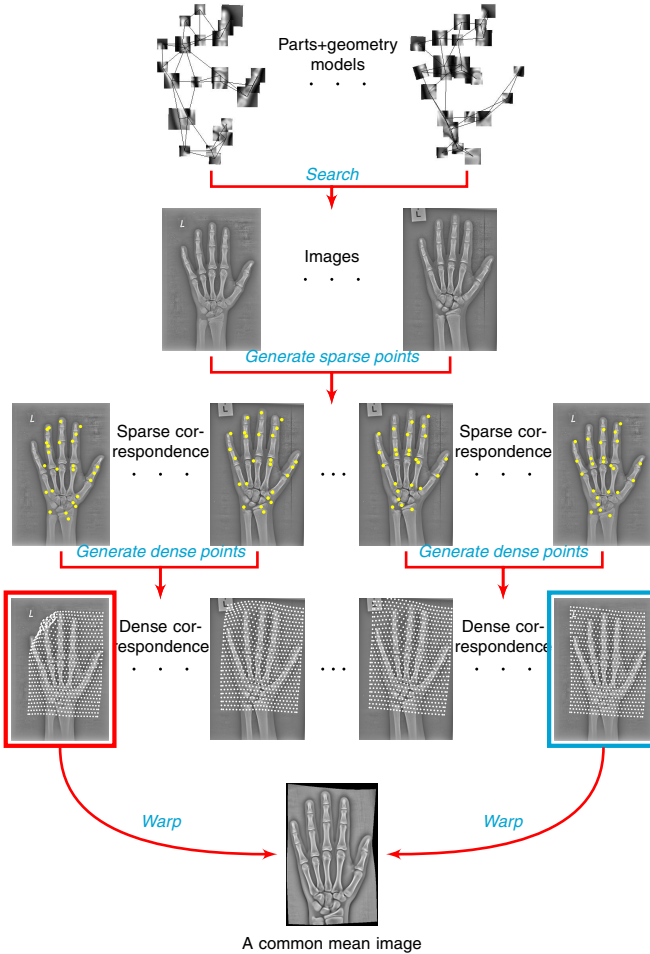
Fig. 2. An overview of the multi-model initialisation strategy

An overview of the multi-model initialisation scheme is given in Fig. 2. Given a set of parts+geometry models, we use each in turn to search the image set to obtain a sparse correspondence, where a dense correspondence can be generated using a thin-plate spline (TPS) interpolation. Poor patterns will lead to poor dense points (indicated by the red box in Fig. 2). The quality of the pattern can be evaluated by warping the target image to a mean image using the dense points, and comparing the similarity between the warped image and the mean. To choose the best pattern for an image, we use each of the associated sets of dense points to warp the image to the mean and compute the similarity. A new correspondence can thus be established by grouping the sets of dense points related to the best patterns across the images. We give the details of each step below.

2.1   Dense Points

To easily compare the quality of different patterns for an image, we transfer the corre-spondence information encoded in each pattern to a common set of points. A simple method is to generate a dense set of points on a reference image and propagate the points to the other images using a TPS. We use $\mathbf{X} = (x_1, y_1, x_2, y_2, \ldots, x_n, y_n)^{\mathrm{T}}$ to denote the positions of the dense points.

2.2   Quality of the Pattern

An observation from groupwise registration is that if the correspondence across the im-age set is well established, we should be able to obtain a crisp mean image. Furthermore, if a pattern is good, its related image should be similar to the mean when comparing the two in the same frame, and *vice versa*. Hence, we can use a mean image to evaluate the quality of a pattern.

Suppose we have a set of images $\{I_k | k = 1, \ldots, N_I\}$ and N_G parts+geometry mod-els. We use $\{\mathbf{X}_k^l | l = 1, \ldots, N_G\}$ to denote the sets of dense points associated with image I_k. Let $\bar{I}$ be the mean image and $\overline{\mathbf{X}}$ be the dense points in the mean (see below). Given an image I_k and one of its dense point sets $\mathbf{X}_k^l$, a warp from $\bar{I}$ to I_k is uniquely defined by $\overline{\mathbf{X}}$ and $\mathbf{X}_k^l$. We write this warp as $\mathbf{z}' = W(\mathbf{z} : \overline{\mathbf{X}}; \mathbf{X}_k^l)$, where $\mathbf{z}$ is a point in $\bar{I}$ and $\mathbf{z}'$ is the corresponding point in I_k. To evaluate the quality of the pattern related to $\mathbf{X}_k^l$, we warp I_k onto $\bar{I}$ so as to compare them in a same frame, and use the following function

$$D_k^l = \sum_{\mathbf{z} \in R} \left| I_k(W(\mathbf{z} : \overline{\mathbf{X}}; \mathbf{X}_k^l)) - \bar{I}(\mathbf{z}) \right|, \qquad (1)$$

where R is a region of interest in the mean frame. This function computes the absolute intensity difference over the region of interest between the warped image $I_k(W(\mathbf{z} : \overline{\mathbf{X}}; \mathbf{X}_k^l))$ and the mean. Note that all the images have been preprocessed to standardise their intensity ranges.

2.3   Mean Image

Given $\{\mathbf{X}_k^l | k = 1, \ldots, N_I\}$, we can perform the following steps to compute the mean image $\bar{I}^l$ and its associated dense points $\overline{\mathbf{X}}^l$:

(1) Align each $\mathbf{X}_k^l$ to a reference frame[1] using Procrustes Analysis [1]. Averaging the aligned dense sets of points leads to $\overline{\mathbf{X}}^l$;
(2) Create a triangulation of $\overline{\mathbf{X}}^l$ using the Delaunay algorithm;
(3) Warp each image I_k to the reference frame, computing $I_k(W(\mathbf{z} : \overline{\mathbf{X}}^l; \mathbf{X}_k^l))$[2] $\mathbf{z} \in R$;
(4) Compute the mean image using $\bar{I}^l(\mathbf{z}) = \frac{1}{N_I} \sum_{k=1}^{N_I} I_k(W(\mathbf{z} : \overline{\mathbf{X}}^l; \mathbf{X}_k^l)), \mathbf{z} \in R$.

---

[1] The choice of the reference frame is free. Any image in the set can be used for this purpose.
[2] $I_k(W(\mathbf{z} : \overline{\mathbf{X}}^l; \mathbf{X}_k^l))$ is computed by piece-wise linear interpolation between corresponding triangles in $\overline{\mathbf{X}}^l$ and $\mathbf{X}_k^l$. We use piece-wise linear interpolation for efficiency.

A Common Mean. Different sets of points $\{\mathbf{X}_k^l | k = 1, \ldots, N_I\}$ will lead to a different mean image $\bar{I}^l$. Hence, the quality D computed using different means cannot be compared directly. This can be solved by using a single, common mean. We take the following steps to compute this common mean image:

(1) For $l \leftarrow 1$ to N_G
    (a) Compute the mean image $\bar{I}^l$ using the sets of points $\{\mathbf{X}_k^l | k = 1, \ldots, N_I\}$;
    (b) Warp every image to the mean $\bar{I}^l$, computing D_k^l using (1);
    (c) Rank all images by D_k^l and select the top 50% of the images as the set $\mathcal{S}^l$;
(2) Find a common set of images $\mathcal{S} = \bigcap_{l=1}^{N_G} \mathcal{S}^l$—selecting the images in all sets $\{\mathcal{S}^l\}$;
(3) For each image in $\mathcal{S}$, average associated sets of dense points, $\widehat{\mathbf{X}}_k = \frac{1}{N_G} \sum_{l=1}^{N_G} \mathbf{X}_k^l$;
(4) Compute a mean image $\bar{I}$ using $\mathcal{S}$ and the sets of points $\{\widehat{\mathbf{X}}_k\}$ (similar to what has been described above). Take this as the common mean.

2.4   Pattern Selection

Once we have the common mean image, we can take the following steps to select the best pattern for each image:

(1) For $l \leftarrow 1$ to N_G
    For $k \leftarrow 1$ to N_I
        Warp I_k to the common mean and compute D_k^l;
(2) For $k \leftarrow 1$ to N_I
    Select the best $\mathbf{X}_k^{\hat{l}}$ from $\{\mathbf{X}_k^l | l = 1, \ldots, N_G\}$ such that $D_k^{\hat{l}}$ is minimum.

By grouping $\mathbf{X}_k^{\hat{l}}$ across the image set we can obtain a new correspondence $\{\mathbf{X}_k^{\hat{l}} | k = 1, \ldots, N_I\}$. Although we can use it to directly initialise groupwise registration, in the following experiment we use it to generate a sparse correspondence. This is more efficient, and allows fairer comparison with the single-model scheme which only outputs sparse sets of points.

To create the sparse correspondence we (1) generate a sparse set of points on an image I_k using the best parts+geometry model (in terms of model utility, see [6] for details); (2) project the sparse points onto the other images using the piece-wise affine transformation between different $\mathbf{X}_k^{\hat{l}}$.

3   Experiments

We demonstrate the approach on three different datasets of increasing difficulty:

(1) 100 digital micrographs of female fly wings. Each image has a size of 1280×1022 pixels and is marked with 15 points by human expert (Fig. 3a);
(2) 100 radiographs of the hands of children (aged between 10-13), taken as part of study into bone ageing. The image size varies across the set, with a height ranging from 1000 to 1700 pixels. Each image has 37 manually labelled landmarks. The resolution of this set of images is 0.2mm per pixel (Fig. 3b);

(a) Fly wing   (b) Hand   (c) Spine
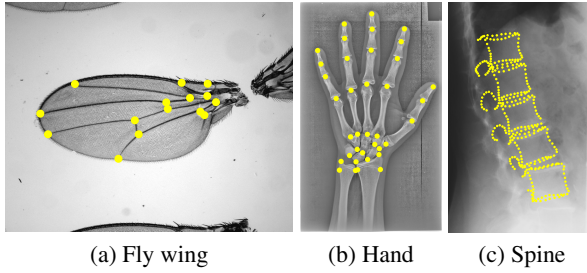
Fig. 3. Examples of the datasets and associated landmarks used in the experiment

(3) 100 radiographs of the lumber spine. The image height varies from 1500 to 1700 pixels. Each image has 337 manual landmarks placed around the outline of the vertebrae. This set of images has a resolution of 0.17mm per pixel (Fig. 3c).

Examples of these three datasets and their manual landmarks are given in Fig. 3. In all experiments the manual annotations are only used to evaluate the performance of the method.

We compare the method with the two single-model initialisation strategies proposed in [6] and [7]. Given a large number of candidate parts, one approach constructs many different parts+geometry models of fixed number of parts, and uses a variant of Genetic Algorithm (GA) to select the best model [6], while the other builds a number of models of random configuration and adopts a voting strategy to achieve the same goal [7]. To generate the candidate parts, both approaches randomly select a reference image, place an overlapping grid on the reference and build a part model for each patch in the grid. In this way, we constructed over 1000, 2700 and 600 candidate parts for the fly wings, hands and spines, respectively. For each set of candidate parts, we first ran the voting based method to choose the optimal parts+geometry model and used it for initialisation. The resulting model determines the number of parts to be used in the GA based approach, from which we obtained a group of N_G models. The very best one was used to do the single-model initialisation as done in [6] and all of them were used for multi-model initialisation. In all experiments we set $N_G = 10$.

To evaluate the accuracy of the groupwise registration we compare with a manual annotation. We used the same protocol with [6] to compute the registration error. Point-to-point location error is reported for the fly wings and hands, and point-to-curve location error is reported for the spines. Results are given in Table 1. We used Welch's (two-tailed) t-test to compare the multi-model strategy with each single-model one, resulting in the p-values. The multi-model initialisation scheme performs similarly to the single-model ones on the fly wings, and does much better on the hands and spines. The power of multiple models can be clearly seen from the result on the spines, where the improvement is the most significant compared with what has been achieved on the other two datasets. This is not surprising as the spine dataset contains large shape variations and fractured vertebrae, which are very challenging for a single model to handle.

To further demonstrate the performance of the multi-model scheme, we repeated the above experiment with another 9 different reference images, thus totally 10 different

Table 1. Point location errors of the dense correspondence

(a) Fly wings (pixels)

Method	Parts	Mean±s.e.	Med.	90%	p-value
GA	20	1.9±0.06	1.7	2.7	$p < 0.01$
Voting	20	1.3±0.05	1.2	1.8	$p = 1$
Multi	20	1.3±0.04	1.2	1.7	–

(b) Hands (mm)

Method	Parts	Mean±s.e.	Med.	90%	p-value
GA	45	1.4±0.1	1.0	2.1	$p < 0.01$
Voting	45	1.3±0.09	1.0	2.2	$p < 0.05$
Multi	45	**1.1±0.04**	**1.0**	**1.8**	–

(c) Spines (mm)

Method	Parts	Mean±s.e.	Med.	90%	p-value
GA	21	3.9±0.5	2.6	4.7	$p < 0.01$
Voting	21	3.6±0.4	2.3	6.5	$p < 0.01$
Multi	21	**2.1±0.2**	**1.6**	**3.1**	–

Table 2. The influence of choice of reference images on the single-model and multi-model initialisation strategies

(a) Fly wings (pixels)

Method	$\hat{c}$	MAD	c_{min}	s.d.
GA	1.8	0.3	1.3	0.5
Voting	1.2	0.09	1.1	0.2
Multi	1.4	0.1	1.2	0.4

(b) Hands (mm)

Method	$\hat{c}$	MAD	c_{min}	s.d.
GA	0.9	0.07	0.8	0.5
Voting	0.9	0.06	0.8	0.6
Multi	0.9	0.04	0.9	0.3

(c) Spines (mm)

Method	$\hat{c}$	MAD	c_{min}	s.d.
GA	2.5	0.6	1.8	6.0
Voting	5.0	2.5	2.3	4.0
Multi	1.8	0.4	1.6	2.0

sets of experiments. This is to show the effects of different choice of the reference image. We computed the median $\hat{c}$ of the 10 medians c_i and the mean absolute difference MAD $= \sum_{i=1}^{10} |\hat{c} - c_i|/10$ for each method on each dataset. We compared $\hat{c}$ with the best median c_{min} of the 10 cases and MAD with the corresponding standard deviation. We summarise the results in Table 2. We find that the choice of reference images only has a small effect on the results of the multi-model initialisation strategy for all three datasets. Although a similar pattern can be observed for both single-model initialisation methods on the fly wings and hands, the performance of the two methods varies more on the spines. For example, the performance of the voting based method varies dramatically from one reference image to another.

4   Discussion and Conclusions

We have described a method that can effectively initialise groupwise non-rigid registration. This is achieved by using multiple parts+geometry models. Experiments show

that our algorithm is able to achieve state-of-the-art results, significantly outperforming earlier approaches that only use a single model. We also compared this multi-model initialisation strategy with the single-model one in terms of the influence of reference images. We find that the multi-model scheme is the least sensitive to the choice of reference images, suggesting that a robust system can be expected.

Current work indiscriminately uses the top models output by the GA based method [6]. If some models result in too many poor matches, the performance of the multi-model scheme will be inevitably degraded. Moreover, if different models share too many common parts, redundancy will arise. This will dilute the advantage of using multiple models to do initialisation, since different models may fail on the same images so that there is no chance to rectify those failures. In the future we will explore how to effectively choose a good set of parts+geometry models. The approach has a natural extension to 3D, which we intend to investigate.

References

1. 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)
2. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active appearance models. IEEE Transactions on Pattern Analysis and Machine Intelligence 23(6), 681–685 (2001)
3. Frangi, A.F., Rueckert, D., Schnabel, J.A., Niessen, W.J.: Automatic 3D ASM Construction via Atlas-Based Landmarking and Volumetric Elastic Registration. In: Insana, M.F., Leahy, R.M. (eds.) IPMI 2001. LNCS, vol. 2082, pp. 78–91. Springer, Heidelberg (2001)
4. Joshi, S., Davis, B., Jomier, M., Gerig, G.: Unbiased diffeomorphic atlas construction for computational anatomy. NeuroImage 23(suppl.1), 151–160 (2004)
5. Cootes, T.F., Twining, C.J., Petrović, V., Babalola, K.O., Taylor, C.J.: Computing accurate correspondences across groups of images. IEEE Transactions on Pattern Analysis and Machine Intelligence 32(11), 1994–2005 (2010)
6. Zhang, P., Adeshina, S.A., Cootes, T.F.: Automatic Learning Sparse Correspondences for Initialising Groupwise Registration. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part II. LNCS, vol. 6362, pp. 635–642. Springer, Heidelberg (2010)
7. Zhang, P., Cootes, T.F.: Automatic Part Selection for Groupwise Registration. In: Székely, G., Hahn, H.K. (eds.) IPMI 2011. LNCS, vol. 6801, pp. 636–647. Springer, Heidelberg (2011)
8. Langs, G., Donner, R., Peloschek, P., Bischof, H.: Robust Autonomous Model Learning from 2D and 3D Data Sets. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part I. LNCS, vol. 4791, pp. 968–976. Springer, Heidelberg (2007)
9. Adeshina, S.A., Cootes, T.F.: Constructing part-based models for groupwise registration. In: Proceedings of International Symposium on Biomedical Imaging, pp. 1073–1076 (2010)
10. Fischler, M.A., Elschlager, R.A.: The representation and matching of pictorial structures. IEEE Transactions on Computer 22(1), 67–92 (1973)
11. Felzenszwalb, P.F., Huttenlocher, D.P.: Pictorial structures for object recognition. International Journal of Computer Vision 61(1), 55–79 (2005)
12. Donner, R., Micusik, B., Langs, G., Bischof, H.: Sparse MRF appearance models for fast anatomical structure localisation. In: Bhalerao, A., Rajpoot, N. (eds.) Proceedings of British Machine Vision Conference, vol. 2, pp. 1080–1089 (2007)
13. Fergus, R., Perona, P., Zisserman, A.: Weakly supervised scale-invariant learning of models for visual recognition. International Journal of Computer Vision 71, 273–303 (2007)

An Efficient and Robust Algorithm for Parallel Groupwise Registration of Bone Surfaces

Martijn van de Giessen[1,2,3,4], Frans M. Vos[1,5], Cornelis A. Grimbergen[4], Lucas J. van Vliet[1], and Geert J. Streekstra[4]

[1] Quantitative Imaging Group, Delft University of Technology, The Netherlands
[2] Division of Image Processing, Leiden University Medical Center, The Netherlands
[3] Department of Intelligent Systems, Delft University of Technology, The Netherlands
[4] Dept. of Biomed. Engineering and Physics, AMC Amsterdam, The Netherlands
[5] Dept. of Radiology, AMC Amsterdam, The Netherlands
`m.vandegiessen@lumc.nl`

Abstract. In this paper a novel groupwise registration algorithm is proposed for the unbiased registration of a *large* number of densely sampled point clouds. The method fits an evolving mean shape to each of the example point clouds thereby minimizing the total deformation. The registration algorithm alternates between a computationally expensive, but parallelizable, deformation step of the mean shape to each example shape and a very inexpensive step updating the mean shape.

The algorithm is evaluated by comparing it to a state of the art registration algorithm [5]. Bone surfaces of wrists, segmented from CT data with a voxel size of $0.3 \times 0.3 \times 0.3 \, \text{mm}^3$, serve as an example test set. The negligible bias and registration error of about 0.12 mm for the proposed algorithm are similar to those in [5]. However, current *point cloud* registration algorithms usually have computational and memory costs that increase quadratically with the number of point clouds, whereas the proposed algorithm has linearly increasing costs, allowing the registration of a much larger number of shapes: 48 versus 8, on the hardware used.

1 Introduction

Groupwise registration is a recurring problem in many medical applications. Two prominent applications are atlas building and the construction of statistical shape models (SSM). Such registrations should be unbiased in the sense that the outcome must not depend on the selection of a target or on the order in which the samples are processed. Furthermore, it is often desirable to register a large number of samples, such that the atlas or SSM generalizes well. However, depending on chosen similarity criteria and allowable transformations the groupwise registration problem may become intractable, both in terms of computational expense as well as memory requirements.

Group-wise registration algorithms are available both for voxel-based registrations, e.g. [3] as for point cloud registrations, e.g. [2]. The most important methodological difference between voxel-based and point cloud registrations is in the correspondence measure, e.g. intensity based vs. distance based. In [6]

it is proposed to approximate the point clouds with Gaussian kernels and to (densely) sample the space with clouds on a grid, effectively transforming the point cloud registration problem into an intensity registration problem. However, using the L_2 divergence and Gaussian kernels to estimate the density function such a sampling can be avoided and the divergence can be computed efficiently in closed form [5]. Unfortunately, the computational and memory costs of the solution proposed in [5] grow quadratically with the number of registered shapes.

In this work we propose a solution to the groupwise point cloud registration problem that has a computational complexity that increases *linearly* with the number of point clouds to be registered. In this algorithm (I) the computational and memory costs grow linearly with the number of clouds, and (II) the algorithm is trivially parallellizable. This allows the unbiased registration of a very large number of point clouds on regular hardware and permits GRID computing. The registration problem is solved by independently evolving copies of a mean cloud that has minimal deformation with respect to each of the point clouds. This solution differs from common approaches where all example point clouds are deformed to the evolving mean [3,6,5] and has the advantage that an implicit point correspondence is present between all registered shapes, without the need for an image grid as in [3,6] This implicit correspondence allows the computationally inexpensive mean shape update. Both the accuracy and efficiency of the proposed method are compared to a state-of-the-art method [5] by applying the registration to three sets of 48 wrist bones (See Figure 3).

2   Methods

The registration algorithm that is proposed in this paper establishes correspondence between N point clouds $C_i, i = 1, \ldots, N$ and an evolving mean cloud M with n_m points. The numbers of points n_i in all clouds C_i do not need to be the same. For each cloud C_i, a deformed copy of M that approximates C_i is denoted as M_i. The algorithm to evolve clouds M and $M_i, i = 1, \ldots, N$ consists of five steps, outlined below. The first three steps are illustrated in Figure 1. After estimating an initial cloud M (step 1), the non-rigid registration in step 2 is the computationally most expensive step. Due to the splitting up of the procedure in a registration (step 2) and update of the mean shape (step 3), step 2 can be performed separately for each cloud C_i and is therefore trivially parallelizable.

Step 1: Estimate an Initial Mean Cloud M. In this work an initial *coarse* alignment of the point clouds C_i is assumed. Furthermore, the initial mean cloud M is assumed to come from a surface with the same topology as the clouds C_i. In this work the surfaces sampled by C_i are available and the initial estimate of M is obtained by sampling the 0-level of the average signed distance transforms of the surfaces.

Step 2: Register M to Each Cloud C_i. Following many recent registration methods, shapes, initially represented as point clouds, are modeled using a

(a) Step 1   (b) Step 2   (c) Step 3
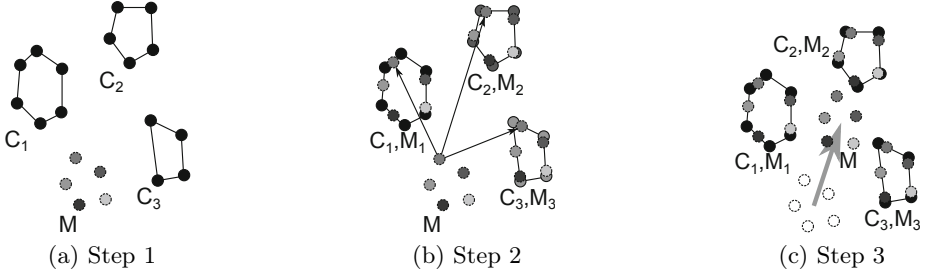
Fig. 1. Schematic representation of the registration algorithm. Step 1: (a) Three point clouds $C_{1,2,3}$ (solid contours) and initial estimate of mean shape M (dashed contours). Step 2: (b) Copies $M_{1,2,3}$ of M have been registered to $C_{1,2,3}$. For one point correspondences are denoted by arrows. Step 3: (c) New estimate of shape M, with minimal deformation with respect to clouds $M_{1,2,3}$.

mixture of Gaussians. In this work, all Gaussian kernels are isotropic and have the same size, determined by parameter σ. The density D_i of each cloud at coordinates $\mathbf{x}$ is therefore described by

$$D_i\left(\mathbf{x}\right) = \frac{1}{n_i\left(2\pi\right)^{d/2}\sigma^d}\sum_{j=1}^{n_i}\exp\left(-\left(\mathbf{x}-\mathbf{p}_{ij}\right)^T\left(\mathbf{x}-\mathbf{p}_{ij}\right)\big/2\sigma^2\right) \tag{1}$$

where n_i is the number of points in cloud C_i, d is the spatial dimensionality of the cloud points and $\mathbf{p}_{ij}$ are the coordinates of a point indexed by j in cloud C_i.

As in [5] the L_2 divergence is used as a distance measure between two density functions. For two clouds M_i and C_i with density functions D_m and D_c, this measure is defined as

$$f_{L2}\left(M_i, C_i\right) = \int_{\mathbb{R}^d}\left(D_m^2 - 2D_mD_c + D_c^2\right)d\mathbf{x} \tag{2}$$

where $\mathbb{R}^d$ is the spatial domain in which the point-clouds reside. The L_2 divergence is a member of the family of Density Power Divergences [5], which also contains the well-known Kullback-Leibler (KL) divergence. The L_2 divergence is symmetric and (2) can be evaluated in closed form for Gaussian density functions, using the identity:

$$\int_{\mathbb{R}^d} G\left(\mathbf{x}|\mu_1, \Sigma_1\right) G\left(\mathbf{x}|\mu_2, \Sigma_2\right)d\mathbf{x} = G\left(0|\mu_1 - \mu_2, \Sigma_1 + \Sigma_2\right) \tag{3}$$

where $G\left(\mathbf{x}|\mu_1, \Sigma_1\right)$ and $G\left(\mathbf{x}|\mu_2, \Sigma_2\right)$ are (multivariate) Gaussian density functions with means μ_1 and μ_2 and covariance matrices Σ_1 and Σ_2, respectively.

In this work the similarity between M_i and C_i is maximized through the minimization of (2). To this end M is transformed (into M_i) with a thin-plate-spline (TPS) transform [1] with n_ϕ control points $\mathbf{p}_l^\phi, l = 1, \ldots, n_\phi$. For cloud C_i

the optimally deformed mean cloud is given by $M_i = \Phi\Theta_i + M$ where Φ is the $n_m \times n_\phi$ matrix that contains the radial basis functions of the TPS transform and Θ_i is the $n_\phi \times d$ matrix that contains the transformation coefficients. The transformation is regularized with the costs for the deformation of M_i

$$f_{stress}\left(\Theta_i, M\right) = \mathrm{tr}\left(\Theta_i^T \Phi^T \Phi\Theta_i\right) \tag{4}$$

where tr stands for the matrix trace. This regularization prevents large deformations from M and thereby preserves shape similarity and meaningful point-point correspondence between the clouds M_i. f_{stress} is a function of M because matrix Φ is a function of M. For all clouds combined, the function to be optimized is defined as the sum

$$F\left(\Theta_1, \ldots, \Theta_N, M\right) = \sum_{i=1}^{N}\left[f_{L2}\left(M_i, C_i\right) + \lambda f_{stress}\left(\Theta_i, M\right)\right] \equiv F_{L2} + \lambda F_{stress} \tag{5}$$

where λ is a regularization weight and $f_{L2}\left(M_i, C_i\right)$ is a function of M and Θ_i, because M_i is a function of M and Θ_i. This function can be minimized directly using iterative methods, such as a quasi-Newton optimization. However, this is a costly optimization, due to the large number of parameters $(n_\phi \cdot d \cdot N)$ and the necessary update of matrix Φ as a function of M. However, by keeping M constant, the cost functions within the sum have no shared optimization variables and can be minimized separately. This the key novelty of our work and of crucial importance to subsequence parallelization.

To account for misalignments, f_{L2} will also include rigid transformations for all clouds C_i. Because the clouds M_i do not deform (only move) with respect to M and f_{stress} can be kept as in (4).

Step 3: Update the Current Estimate of M by Computing the Mean of M_i. Keeping M constant during the minimization of (5) prohibits finding the global minimum of (5). Therefore M needs to be updated separately. With M_i constant, M only affects the term that describes the total deformation costs $F_{stress} = \sum_{i=1}^{N} \mathrm{tr}\left(\Theta_i^T \Phi^T \Phi\Theta_i\right)$. From $\Phi\Theta_i = M_i - M$ follows that

$$F_{stress} = \sum_{i=1}^{N} \mathrm{tr}\left((M_i - M)^T (M_i - M)\right) = \sum_{i=1}^{N}\sum_{j=1}^{n_m} \left\|\mathbf{p}_j^{m_i} - \mathbf{p}_j^m\right\|^2$$

where $\mathbf{p}_j^{m_i}$ and $\mathbf{p}_j^m$ are the points with index j in cloud M_i and M, respectively. Therefore, F_{stress} is minimal when $\mathbf{p}_j^m$ is the mean of $\mathbf{p}_j^{m_i}, i = 1, \ldots, N$, thus the optimal estimate of the mean shape is given by $M \leftarrow \frac{1}{N}\sum_{i=1}^{N} M_i$. This inexpensive step takes care of the coupling of the clouds. Note that this simple mean computation is only possible because of the implicit correspondence between all deformed mean shapes, which is particular for the proposed algorithm. Θ_i is updated using a linear least-squares estimate.

Step 4: Test for Convergence. If converged, continue, otherwise go to step 2.

Step 5: Transform C_i to the Mean Cloud Using the Correspondence between M_i and M. One should note that minimizing $F\left(\tilde{\Theta}_1, \ldots, \tilde{\Theta}_N, M\right)$ as in (5) results in sets of corresponding points M_i. However, the much denser clouds C_i have not deformed and, thus, are not registered to M (See Figure 1c). To do so, all point clouds C_i are rigidly transformed using $\mathbf{r}_i$ and $\mathbf{t}_i$ and deformed towards M by computing the inverse TPS transforms $C_i = \Phi_i^C \Theta_i + \hat{C}_i$ where the matrix with TPS basis functions Φ_i^C and the deformed point cloud $\hat{C}_i$ have to be estimated for a given set of transformation parameters Θ_i. This can be done using an iterative procedure. For an accurate registration, all clouds $\hat{C}_i$ represent a surface of the same shape. These registered surfaces can then be used for a *dense* correspondence estimate between the clouds C_i.

3   Experiments

The experiments in this section evaluate both the accuracy and the precision of the proposed registration algorithm and compare these to the registration method in [5]. The chosen application is the registration of wrist bones to establish a dense correspondence of points on the bone surfaces of different individuals. Specifically, the focus is on the scaphoid, lunate and hamate bones (See Figure 3). The 3×48 bone surfaces are represented as triangulated surfaces with vertices V_i at a sampling density of approximately 0.14 vertices/mm^2 (voxel size $0.3 \times 0.3 \times 0.3$ mm^3), typically yielding 1.8×10^4 (SD 3.4×10^3) points per bone surface. The bones are coarsely aligned by ensuring the same scan orientations and by translating the centers of gravity of each bone to the origin. Optimal parameter settings (See Section 2) were experimentally determined for this data as $\sigma = 0.6$ mm, $n_m = 1000$, $\lambda = 10^{-6}$ and $n_\phi = 600$. Each bone is represented by a point cloud C_i with a subset of n_m vertices of V_i.

In all experiments the registration accuracy E_{acc} and precision E_{prec} are evaluated on the transformed bone surfaces i.e. after the vertices V_i are deformed towards M as in Step 5. This allows comparable results for different sample densities. The resulting clouds are denoted as $\hat{V}_i$. E_{acc} is defined as the average (over shapes) norm of the average (over points) *signed* point-to-plane distance between all pairs of clouds $\hat{V}_i$ and $\hat{V}_j$, evaluated on points that are not in C_i (thus not used for the registration). E_{prec} is the average *absolute* point-to-plane distance between all pairs of clouds $\hat{V}_i$ and $\hat{V}_j$, again using points not in C_i. E_{acc} reflects the presence of a bias, while E_{prec} reflects the remaining matching error.

3.1   Robustness to Initial Mean Cloud Estimate

The initial estimate of M has a strong influence on F_{stress} during the first iterations. The robustness of the algorithm was tested by using four different initial estimates of M: (I) Points drawn randomly from the 0-level set of the average signed distance transform (SDT) as in Section 2, Step 1. (II) Points drawn randomly from all clouds to be registered (rand). (III) Points drawn
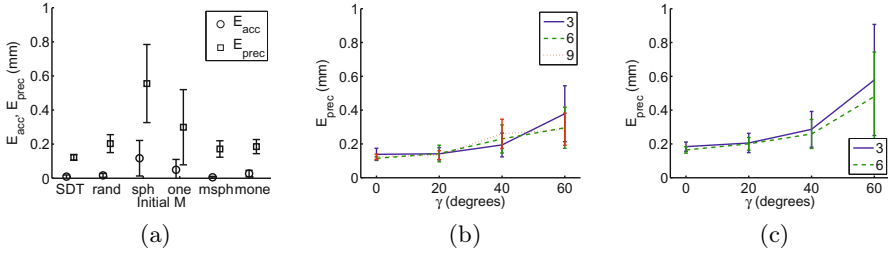
Fig. 2. (a) Registration accuracy and precision for four initial estimates of the mean shape M: (I) SDT (II) rand, (III) sph, (IV) one. For 'sph' and 'one' also the results of a multiscale registration are shown (msph, mone). (b) Registration precision for the proposed algorithm and (c) for the algorithm from [5] for increasing γ and N. (a-c) The errorbars denote standard deviations.

randomly from a sphere with radius equal to the mean standard deviation of the point cloud coordinates (sph). (IV) One of the point clouds (one). For each of these initial point clouds a random selection of $N = 3$ three clouds were registered, repeated with 10 drawings for each of the 3 bone types.

The registration accuracies E_{acc} and precisions E_{prec} in Figure 2a show that initializing M by taking points from the SDT gives the most accurate results, closely followed by a random selection of points from the non-registered point clouds. With a sphere or a single shape the registration algorithm often converged in local minima due to a lack of overlap between the kernels of M and the kernels of C_1, C_2 and C_3 for the current, small, choice of $\sigma = 0.6$ mm. A multiscale approach where σ decreases from $\sigma = 3$ to $\sigma = 0.6$ mm solves this as depicted by the last two results of Figure 2a.

3.2   Robustness to Initial Shape Alignment

When more shapes are present, rotated over a random angle, it is more likely that a shape that has a rotation 'in between' improves the convergence of the algorithm. Therefore N previously aligned shapes were rotated around randomly distributed rotation axes, with angles randomly sampled from $[-\gamma, \gamma]$, for different values of N and γ. The initial M was obtained as in Section 2, step 1. The clouds were registered using both the proposed algorithm and the algorithm in [5]. The experiment was repeated 30 times for each combination of the following settings: $N \in \{3, 6, 9\}$ randomly selected shapes (10 selections for each of the three bone types) and $\gamma \in \{20, 40, 60\}$ degrees.

E_{acc} for the proposed method and the method from [5] were all in the order of 10^{-4} mm, except at $\gamma = 60°$. Here E_{acc} was approximately 0.07 mm for $N = 3$ shapes. In the latter case, both algorithms converged in a local minimum with large shape deformations. E_{prec} is shown in Figure 2b for the proposed method and in Figure 2c for the method in [5]. For $\gamma = 0°, 20°$ both methods perform equally well. For $\gamma = 40°, 60°$ the proposed method is more precise than the
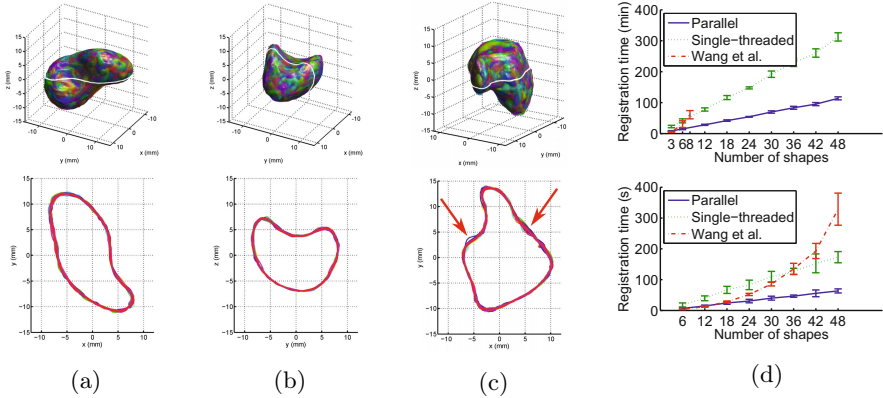
Fig. 3. Bone surfaces and intersections after registration of $N = 48$ (a) scaphoids, (b) lunates and (c) hamates. The white lines on the surfaces show the locations of the contours. (d) Average registration times (and standard deviations) in minutes for the registration of $N \in \{6, 12, \ldots, 48\}$ shapes, for the proposed method parallel on three cores and single-threaded and the method in [5] for $n_m = 1000$ (top) and $n_m = 100$ (bottom).

method of [5]. We hypothesize that the method of [5] is slightly more susceptible to local minima, due to the larger number of concurrently optimized parameters. Furthermore, an increase in N decreases the registration error for large angles γ. This is mainly due to the denser sampling of poses between $-\gamma$ and γ. For [5] experiments with 9 shapes did not succeed due to a lack of computer memory.

3.3   Feasibility of Large Data Set Registration

To investigate the feasibility of registering large datasets, an increasing number of randomly selected sets of bones were registered: $N \in \{6, 12, \ldots, 48\}$. Registrations were performed for each type of bone and for three different shape set selections. Experiments were performed on a computer with an Intel Xeon processor at 2.67GHz with 6.0 GB of RAM memory. The method was implemented in MATLAB R2010b, from The Mathworks, Inc.

Example registration results are shown in Figure 3 for $N = 48$ bones. For the scaphoid and lunate, all contours are aligned. For the hamate, however, small misalignments can be observed on the left and right of the protrusion of the bone (see arrows). This is due to the large shape variations of the protrusion, combined with the TPS interpolation of the shape surfaces. For correspondence estimates between the surfaces, however, these small misalignments do not form a problem. For 12 shapes and more, the registration accuracy E_{acc} and precision E_{prec} do not depend on the number of shapes and are $E_{acc} \approx 0.00$ mm and $E_{prec} \approx 0.12$ mm, with negligible standard deviations.

Figure 3d shows that the registration time increases linearly with an increasing number of shapes, using the proposed method, while it increases approximately quadratically using the method proposed in [5]. For $n_m = 1000$, registering more than 8 shapes, using the method of [5] was not possible, as the data did not fit in the availabe RAM. Therefore also results for $n_m = 100$ are shown. The parallelization of the registration using three cores shortens the registration time with approximately a factor 2.8, an efficient parallelization.

4   Discussion

In this paper, a method was presented for the non-rigid registration of a large number of shapes, whose surfaces are represented by point clouds. Experiments showed that the proposed algorithm indeed can register large numbers of point clouds with high accuracy and precision with modest hardware demands. The registration time increases linearly with the number of shapes N. Furthermore, the registration accuracies and precisions are similar to the method of [5], which itself was compared favorably to other state of the art methods, e.g. [6].

Although, as in [2] both algorithms evolve a mean cloud that is only mildy constrained, the proposed method does not suffer from the instabilities in [2]. This is because the proposed method by definition does not need the assumption that the 'forward' and 'backward' thin-plate spline transform are exactly the same and because of an improved similarity measure. Interesting follow-up research is if current based methods, e.g. [4] also allow group-wise registration with a closed-form mean estimate as in step 3.

As shown in Figure 3, a much denser correspondence than the $n_m = 1000$ points used for registration can easily be obtained from the registered surfaces. Furthermore, taking the linear increase of registration times into account, combined with the parallelization of the non-rigid registration step, one could register almost 600 surfaces in a day. Note that such a dataset is not easily obtained.

References

1. Bookstein, F.L.: Principal warps - thin-plate splines and the decomposition of deformations. IEEE Trans. on Patt. An. and Mach. Intell. 11(6), 567–585 (1989)
2. Chui, H., Rangarajan, A., Zhang, J., Leonard, C.M.: Unsupervised learning of an atlas from unlabeled point-sets. IEEE Trans. on Patt. An. and Mach. Intell. 26(2), 160–172 (2004)
3. Joshi, S., Davis, B., Jomier, M., Gerig, G.: Unbiased diffeomorphic atlas construction for computational anatomy. Neuroimage 23, S151–S160 (2004)
4. Myronenko, A., Song, X.: Point set registration: Coherent point drift. IEEE Trans. on Patt. An. and Mach. Intell. 32(12), 2262–2275 (2010)
5. Wang, F., Vemuri, B., Syeda-Mahmood, T.: Generalized L2-Divergence and Its Application to Shape Alignment. In: Prince, J.L., Pham, D.L., Myers, K.J. (eds.) IPMI 2009. LNCS, vol. 5636, pp. 227–238. Springer, Heidelberg (2009)
6. Wang, F., Vemuri, B.C., Rangarajan, A., Eisenschenk, S.J.: Simultaneous nonrigid registration of multiple point sets and atlas construction. IEEE Trans. on Patt. An. and Mach. Intell. 30(11), 2011–2022 (2008)

Realistic Head Model Design and 3D Brain Imaging of NIRS Signals Using Audio Stimuli on Preterm Neonates for Intra-Ventricular Hemorrhage Diagnosis

Marc Fournier[1], Mahdi Mahmoudzadeh[1], Kamran Kazemi[2], Guy Kongolo[1], Ghislaine Dehaene-Lambertz[3], Reinhard Grebe[1], and Fabrice Wallois[1]

[1] GRAMFC, Inserm U1105, CHU Amiens, University of Picardie Jules Verne, Amiens, France
{marc.fournier,mahdi.mahmoudzadeh,guy.kongolo,reinhard.grebe,
fabrice.wallois}@u-picardie.fr
[2] Dept. of Electrical and Electronics Engineering, Shiraz University of Technology, Shiraz, Iran
kazemi@sutech.ac.ir
[3] NeuroSpin, CEA, Inserm U992, Cognitive Neuroimaging, University Paris XI, Paris, France
ghislaine.dehaene@cea.fr

Abstract. In this paper we propose an auditory stimulation and Near Infra-Red Spectroscopy (NIRS) hemodynamic changes acquisition protocol for preterm neonates. This study is designed to assess the specific characteristics of neurovascular coupling to auditory stimuli in healthy and ill neonate brains. The method could lead to clinical application in Intra-Ventricular Hemorrhage (IVH) diagnosis along with other techniques such as EEG. We propose a realistic head model creation with all useful head structures and brain tissues including the neonate fontanel for more accurate results from NIRS signals modeling. We also design a 3D imaging tool for dynamic mapping and analysis of brain activation onto the cortex surface. Results show significant differences in oxy-hemoglobin between healthy neonates and subjects with IVH.

Keywords: Head model design, 3D brain NIRS optical imaging, Hemodynamic response to audio stimuli, Neonate IVH diagnosis, Functional brain imaging.

1 Introduction

This paper introduces a Near Infra-Red Spectroscopy (NIRS) cerebral hemodynamic response optical monitoring method of the healthy and sick premature infant with Intra-Ventricular Hemorrhage (IVH). Non-invasive investigation of the oxygenation of the infant's brain is of high interest. The proposed method can be applied for early diagnosis of impairments in complement with other current brain imaging techniques such as ElectroEncephaloGraphy (EEG). Low arterial blood oxygenation and abnormal cerebral blood flow is believed to influence the function of the neonatal brain [1]. Preterm neonates are at high-risk of IVH because of their lack of ability to regulate cerebral blood flow and pressure [2]. Currently, brain injuries in infants are mainly diagnosed clinically by EEG (investigates neural function, but not vascular response) and cranial ultrasound (gives only anatomical information) [3]. Previous studies on cerebral hemodynamic

responses using NIRS on neonates have been done using various stimulations such as visual [4] and auditory [5]. Visual stimuli in adults using NIRS has also been proposed while investigating head modeling and image reconstruction [6]. Another study on cerebral blood flow estimated by NIRS in preterm neonates has correlated NIRS findings with cerebral ultrasound results [7]. Children with delays in language and speech development are at high risk for later disorders in reading, spelling, and writing; academic skills which are highly dependent on language abilities. Brain injuries have been found to be associated with language and speech outcomes among children born prematurely [8]. IVH are risk factors for adver-se neurodevelopment outcomes, including cognitive impairment and cerebral palsy.

2   Material and Methods

2.1   Subjects and NIRS Signals Acquisition Setup

This study was carried out on two groups; the first one is composed of 12 healthy control subjects and the second one of 7 ill subjects with IVH of grade III-IV. All subjects of both groups are preterm neonates of gestational age from 28 to 32 weeks; tested during their sleep between 2 and 4 days after birth. Subjects were submitted to auditory stimuli which consist of two digitized syllables /ba/ and /ga/ as in a previous EEG study [9]. Three stimulations conditions were used: the standard one (ST: four /ba/ male); deviant voice (DV: three /ba/ male, one /ba/ female); and deviant phoneme (DP: three /ba/ male, one /ga/ male). The four syllables block duration is 4s and total stimulation (20s) is composed of five consecutive blocks. A newborn special NIRS probe showed in Fig. 1(a) was designed and consists of two patches containing two detectors and sixteen light sources in each of them (8 to λ=690nm; 8 to λ=830nm wavelengths). Twenty acquisition channels, ten per hemisphere, are measured in the configuration showed in Fig. 1(b). We used a multi-channel frequency domain based optical imaging system (Imagent, ISS Inc.) for the acquisition of oxygenated hemoglobin (HbO) and deoxygenated hemoglobin (Hb) changes during auditory stimuli. Values of HbO and Hb and their changes were obtained using the relation between absorption spectroscopic coefficients of the environment and chromophore concentrations according to the modified Beer-Lambert law [10] used in NIRS studies.
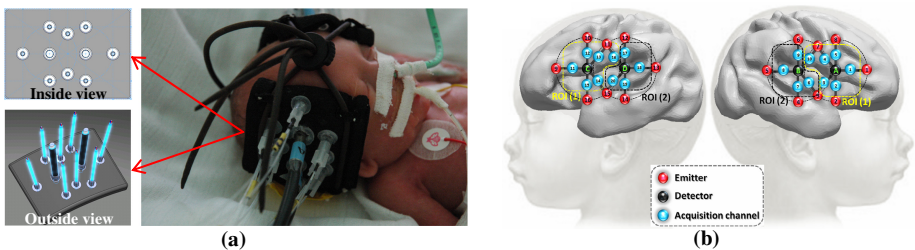

Fig. 1. NIRS acquisition probe design. (a) NIRS sensors patches design and its application to one subject. (b) Schematic positions and numbering of NIRS sensors according to both hemispheres with 4 detectors (A...D, black), 16 emitters (1...16, red) and 20 channels (1...20, blue).

2.2   Neonate Realistic Head Model Creation

We created a 30 weeks (average of all healthy and ill subjects) preterm neonate complete head model with all useful head and brain structures used for NIRS signals mapping onto the brain. The first of three steps was a straightforward extraction from choosing the 30 weeks iteration of a previously created 4D (3D+t) preterm neonate MRI-based atlas [11]. This results in intracranial head model creation including cerebrospinal fluid, grey matter and white matter tissues. In the second step we created a full term neonate realistic head model from 15 subjects between 39 to 42 weeks of gestational age. MRI and CT-Scan of all subjects were acquired for clinical purposes and were found without any anatomical abnormalities after inspection. We used SPM toolbox (http://www.fil.ion.ucl.ac.uk/spm/) to perform co-registration of both imaging modalities, segmentation and space normalization of head and brain structures in order to build the atlas containing each structure probability mask. The fontanel was segmented from the CT-Scan using a variational level-set method [12]. The third step was achieved by registering the intracranial brain mask of the full term head model (created at step 2) to the intracranial brain mask of the preterm head model (created at step 1). The resulting transformation was then applied to the scalp, skull and fontanel of the full term model in order to fit these structures onto the preterm intracranial model; thus obtaining a complete preterm head model with all useful structures. Analysis of our full term neonate head model is resumed in Fig. 2 which shows the importance of considering the fontanel in neonate brain modeling; especially in temporal regions, of interest for this study, that are more widely covered by the fontanel.
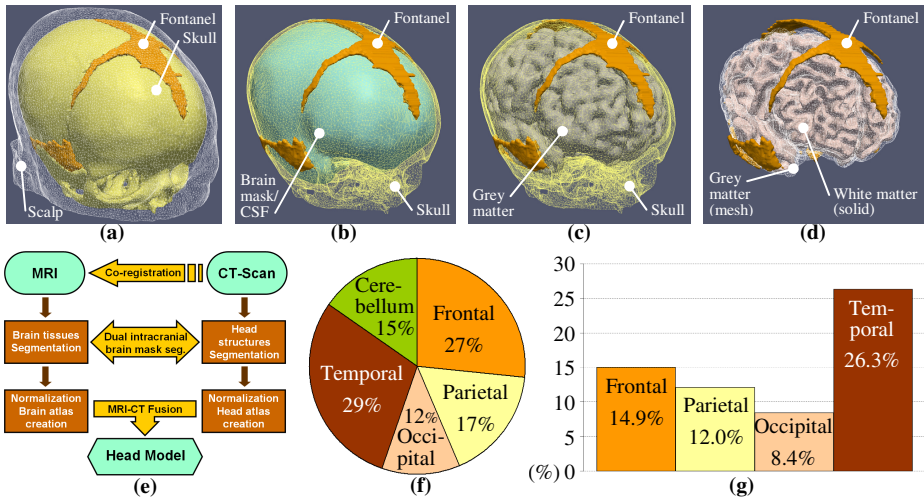
Fig. 2. Impact of fontanel in the head model. (a) to (d) Views of fontanel and other head and brain structures on one subject used in the head model creation. The brain is covered at 17% by the fontanel and 83% by the skull. (e) Procedure to create the full term head model (step two). (f) Proportion of the fontanel overlapping each brain lobes: this means 29% of the total fontanel surface overlaps both brain temporal lobes. (g) For each brain lobe, proportion of the overlapping fontanel compared to the total surface of the lobe: this means 26.3% of both temporal lobes surface is covered by the fontanel. Charts show average values for all subjects.

To create our full term neonate head model we segment brain tissues from the MRI using the expectation maximization method [13]. To overcome the large grey levels overlap between different tissues in neonate MRI, we incorporate contextual and spatial a priori information during tissue classification using a mixture of Gaussians where each cluster is modeled by its mean, variance and a mixing proportion.

Skull and fontanel are segmented from CT-Scan. In neonates, the skull is composed of cranial bones and the fontanel. The cranial bones are obtained by applying automatic threshold to the image histogram and the fontanel can be identified as gaps between cranial bones of the skull. In order to extract the fontanel, a model-based variational level-set method is used to reconstruct the full skull; thus filling the fontanel gaps. Then the fontanel is obtained by removing cranial bones from the reconstructed full skull. Let φ be the level-set function for segmentation and φ_m be the signed distance function of the skull model. To reconstruct the full skull (including fontanel) the input image is segmented by minimizing the following functional:

$$E(c_1,c_2,\varphi) = E_{CV}(c_1,c_2,\varphi) + E_{shape}(\varphi), \quad with: \ E_{shape} = \int_{\Omega} (\varphi(\vec{x}) - \varphi_m(\vec{x}))^2 \, dx, \quad and:$$

$$E_{CV} = \lambda_1 \int_{\Omega} (I(\vec{x}) - c_1)^2 H(\varphi) d\vec{x} + \lambda_2 \int_{\Omega} (I(\vec{x}) - c_2)^2 (1 - H(\varphi)) d\vec{x} + \nu H(\varphi) + \mu |\nabla H(\varphi)|$$

(1)

where E_{CV} is a level-set formulation of the functional as previously proposed in [14] with $H(\varphi)=[1+(2/\pi)\mathrm{atan}(\varphi/\varepsilon)]/2$ being the heaviside function. E_{shape} is used as shape prior to guide the segmentation as detailed in [12]; thus providing a constraint during evolution of the level-set function which is attracted to the skull.

Both MRI and CT atlases are created using the same method based on the intracranial brain mask (*BM*) segmentation of each subject in order to register all other brain tissues and head structures (*ST*) according to this procedure:

1. Select a reference subject: BM_R
2. Compute each subject affine transformation: $A_i = BM_i \rightarrow BM_R$
3. Normalize (first-pass) each subject structures: $ST_i' = A_i \times ST_i$
4. Compute cosine basis functions deformation fields: $D_i = BM_i' \rightarrow BM_R$
5. Normalize (second-pass) each subject structures: $ST_i'' = D_i \times ST_i'$
6. Compute nonlinear transf. of ref. subj. to all affine norm. subj.: $T_i = BM_R \rightarrow BM_i'$
7. Compute the mean transformation: $T_M = \mathrm{mean}(T_i)$
8. Compute spatially unbiased structures: $ST_{Ui} = T_M \times ST_i''$
9. Compute each structure average from individuals: $ST_{UA} = \mathrm{mean}(ST_{Ui})$
10. Smooth ST_{UA} structures with a Gaussian kernel: 2mm full-width half-maximum

To minimize the bias introduced by the chosen reference subject, this atlas creation process is repeated by replacing the initial reference intracranial target (BM_R) with the first-pass intracranial model result. Finally the created MRI and CT atlases are fused together to create the multimodality neonatal atlas. The intracranial probabilistic model created from CT data is registered onto the MRI one. The obtained transformation parameters are applied to the CT atlas head structures in order to complete the MRI atlas brain tissues; thus obtaining our full term neonate head model.

2.3   Modeling and Imaging of NIRS Signals Propagation in Tissues

At NIRS signals acquisition session, the position of each emitter and detector are digitized along with biomarkers reference positions. For each subject these 3D sensors positions are registered onto our average 30 weeks preterm neonate head model. Then the banana-shape photon path [15] through the head and brain tissues between each considered pair of emitter/detector are computed. Fig. 3(a) shows a schematic of the photon path between a pair of emitter/detector. Fig. 3(b) shows one subject example of the 3D head model with the sensors positions and the computed photon paths. The photons migration along the path between the emitter and detector is modeled by a probability distribution [16] which occurs with a hitting density expressed by:

$$P(x, y, z) = \frac{z^2 \exp[-k(\sqrt{A} + \sqrt{B})] \cdot (k\sqrt{A} + 1) \cdot (k\sqrt{B} + 1)}{A^{3/2} \cdot B^{3/2}} \tag{2}$$

$$with: A = x^2 + y^2 + z^2, \ B = (d - x)^2 + y^2 + z^2, \ and: \ k = \sqrt{3\mu_a \mu_s}$$

where d is the emitter/detector distance (15mm), μ_a the absorption coefficient and μ_s the reduced scattering coefficient of the environment. Our anatomical head model is used to determine to which tissue belongs each voxel in order to locally compute the suitable k based on the different values of μ_a and μ_s provided in [17] for each tissue. It has been demonstrated in [17] that considering the fontanel tissue in our head model leads to a more realistic modeling from NIRS signals. For each subject we constraint equation (2) to the grey matter of the preterm head model which is filled with information according to all photon paths and their distributions. Each pair-wise sensor photon path may partially overlap into some voxels; resulting in multiple values for a single voxel. Thus an average function is used to output a final value per voxel. Then a grand average is performed over all subjects and brain activity is normalized. For visualization purpose, brain activations in the grey matter volume are integrated to the cortex surface by orthogonal projection onto the brain folds themselves. This procedure is repeated for each time steps of the recorded NIRS signals. All generated 3D images of brain activations build a real time series of dynamic topography mapping.
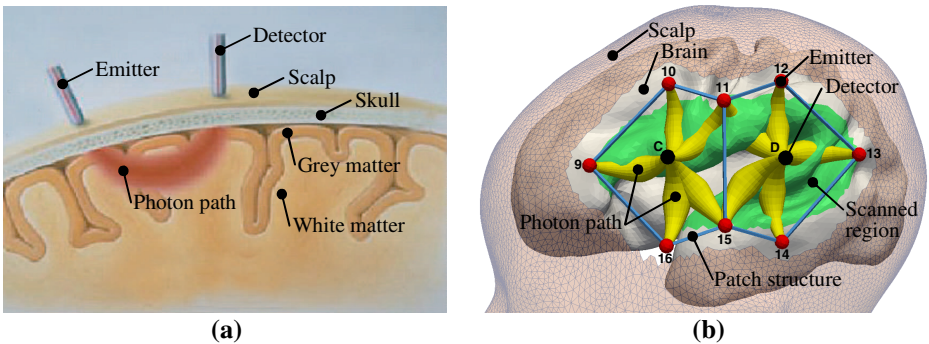

(a)   (b)

Fig. 3. Photon path and its 3D modeling. (a) Schematic of the banana-shape photon path trough head tissues between sensors. (b) Left hemisphere view of the 3D head model (only scalp and brain are showed) with NIRS sensors registered and 3D photon paths trough the grey matter.

3   Results

NIRS signals are acquired at a frequency of 9.2Hz along with oxygen saturation and cardio-respiratory sensors data used in preprocessing to reject artifacts due to systemic and non-cerebral modifications. A software application has been designed to compare in real time healthy and IVH subjects dynamic brain activations images generated as described in previous section 2.3. This tool allows analysis of dynamic mapping of NIRS signals onto the brain surface which runs from -5s (5s before stimuli start, used for baseline information) to +25s (5s after the 20s of stimuli, used for resting state).

Fig. 4(a) is an average image of 5s activity of NIRS HbO signals. It was computed using the dynamic real time brain activation mapping application. The average period from 5s to 10s of stimulation time was chosen for its high and peak activity as shown in most HbO channels in Fig. 4(b). For each stimulation condition (ST, DV, DP) taken individually and for all conditions together, Fig. 4(a) shows higher brain activation in normal subjects compared to abnormal IVH ones.

(a)

(b)

(c)

(d)
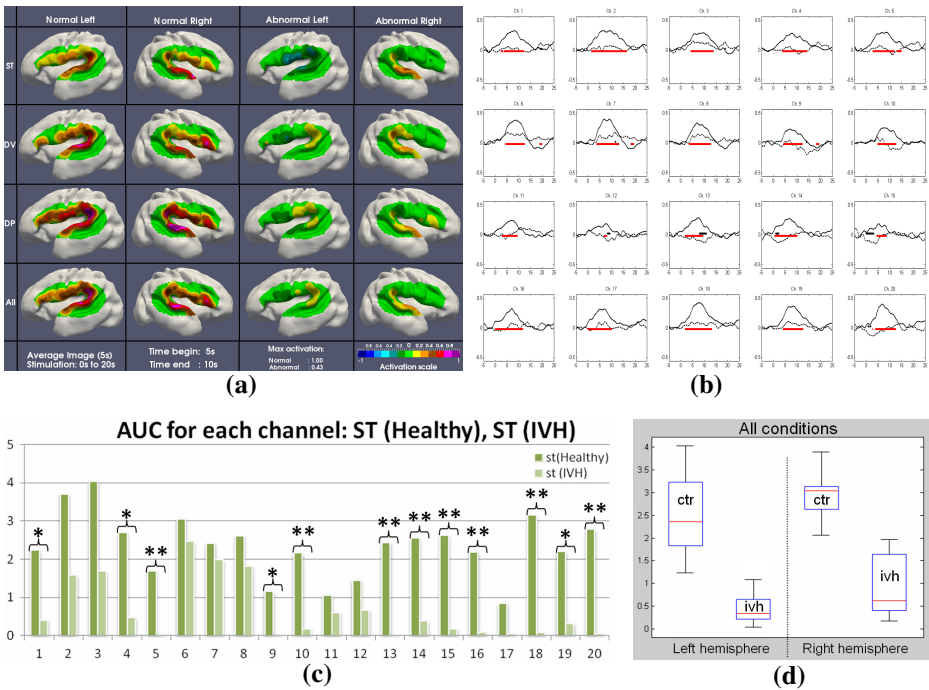
Fig. 4. Images and signals comparisons between healthy and IVH subjects. (a) Average image of 5s (5s to 10s) showing much more activation in healthy subjects for each condition. (b) Grand average HbO signals for the 20 channels over all conditions and significant part of the signals at p<0.05 corrected for multiple comparisons. (c) AUC per channel for ST condition. Channels marked with at least one star (*) show significant t-test difference at p<0.05. Channels marked with two stars (**) are corrected for multiple comparisons. (d) AUC per hemisphere for all conditions showing no overlap between healthy control (ctr) and IVH subjects.

The strongest brain activation shown in Fig. 4(a) for IVH is only 43% of the maximum in normal subjects and it is only visible on the planum temporal of right hemisphere for standard condition. Otherwise the maximum activation of IVH found elsewhere is around 25% of healthy subjects. Normal control subjects show high activation in all conditions and most regions of interest. Fig. 4(b) shows the 20 NIRS channels HbO signals of all conditions averaged together. The solid line is for healthy subjects and the doted line for IVH ones. The gap between the levels of activation of both groups is obvious in all channels. The horizontal red (healthy) and black (IVH) bars show the significant part of the signals compared to the baseline (-5s to 0s) at p<0.05 corrected for multiple comparisons using Holm's method for small samples [18] implemented in FieldTrip toolbox (http://fieldtrip.fcdonders.nl/). The maximum number of permutations for each group has been used. This result shows large significant activations in most channels for healthy subjects and almost no significant activation for IVH subjects. The main significant part of IVH signals are almost exclusively desactivation (negative values) which reinforce the gap between both groups.

Fig. 4(c) shows the area under curve (AUC) integration values for each channel signals in the standard condition for both groups. Standard condition has been selected because it shows less activation in the healthy group. Moreover the AUC was computed for the positive part of the signals only since IVH subjects show more desactivation. These choices have been made to reduce to minimum the gap between both groups. Even in these analysis conditions, Fig. 4(c) result shows statistically significant differences between both groups in the majority of the channels which are star marked (* t-test at p<0.05; ** corrected with Holms method). Fig. 4(d) shows another AUC result for all conditions together and per hemisphere which means average values for the ten channels on each side have been used. This box-plot type result shows no overlap at all between both groups in AUC of all individual subjects; suggesting a strong significant difference between both groups.

4   Conclusion

The present data confirm the existence of neurovascular coupling in premature brain, although it may be impaired in ill subjects. It also shows that ill premature neonates are unable to process syllabic stimuli, a step for language acquisition ability. This inability of pathological brain to adapt to either endogenous or exogenous stimuli by an increase in blood flow can represent a mechanism by which the pathological brain enters in a deleterious pathological loop. This might explain cerebral disabilities observed latter in acquisitions throughout the neurodevelopment [8]. Our audio stimuli and NIRS acquisition protocol coupled to the designed head model and 3D dynamic mapping of brain activation could represent a valuable tool for IVH diagnosis and for a wider range of applications as well. On top of the binary classification of a subject (healthy or IVH), the proposed imaging solution enables: i) a multimodality image-based analysis to compare and complete other sources information such as EEG, ultrasound scan and fMRI; and ii) a shape/functional related local region analysis such as their asymmetries.

References

1. Volpe, J.J.: Brain injury in the premature infant: Neuropathology, clinical aspects, and pathogenesis. Clin. Perinatol. 3, 567–587 (1997)
2. Ballabh, P.: Intraventricular hemorrhage in premature infants: mechanism of disease. Pediatr. Res. 67, 1–8 (2010)
3. Gibson, A.P., Hebden, J.C., Arridge, S.R.: Recent advances in diffuse optical imaging. Phys. Med. Biol. 50, 1–43 (2005)
4. Hoshi, Y., Kohri, S., Matsumoto, Y., Cho, K., Matsuda, T., Okajima, S., Fujimoto, S.: Hemodynamic responses to photic stimulation in neonates. Pediatr. Neurol. 23, 323–327 (2000)
5. Sakatani, K., Chen, S., Lichty, W., Zuo, H., Wang, Y.P.: Cerebral blood oxygenation changes induced by auditory stimulation in newborn infants measured by near infrared spectroscopy. Early Hum. Dev. 55, 229–236 (1999)
6. Dehghani, H., White, B.R., Zeff, B.W., Tizzard, A., Culver, J.P.: Depth sensitivity and image reconstruction analysis of dense imaging arrays for mapping brain function with diffuse optical tomography. Applied Optics 48, D137–D143 (2009)
7. Meek, J.H., Tyszczuk, L., Elwell, C.E., Wyatt, J.S.: Low cerebral blood flow is a risk factor for severe intraventricular haemorrhage. Arch. Dis. Child Fetal Neonatal Ed. 81, 15–18 (1999)
8. Downie, A.L., Frisk, V., Jakobson, L.S.: The impact of periventricular brain injury on reading and spelling abilities in the late elementary and adolescent years. Child Neuropsychol. 11, 479–495 (2005)
9. Dehaene-Lambertz, G.: Cerebral specialization for speech and non-speech stimuli in infants. J. Cogn. Neurosci. 12, 449–460 (2000)
10. Delpy, D.T., Cope, M., Van-Der-Zee, P., Arridge, S., Wray, S., Wyatt, J.: Estimation of optical pathlength through tissue from direct time of flight measurement. Phys. Med. Biol. 33, 1433–1442 (1988)
11. Murgasova, M., Aljabar, P., Srinivasan, L., Counsell, S.J., Doria, V., Serag, A., Gousias, I.S., Boardman, J.P., Rutherford, M.A., Edwards, A.D., Hajnal, J.V., Rueckert, D.: A dynamic 4D probabilistic atlas of the developing brain. Neuroimage 54, 2750–2763 (2011)
12. Jafarian, N., Kazemi, K., Abrishami-Moghaddam, H., Grebe, R., Fournier, M., Helfroush, M.S., Gondry-Jouet, C., Wallois, F.: Automatic segmentation of newborns' skull and fontanel from CT data using model-based variational level set. Signal Image and Video Processing (to appear, 2012), doi:10.1007/s11760-012-0300-x
13. Ashburner, J., Friston, K.J.: Unified segmentation. Neuroimage 26, 839–851 (2005)
14. Chan, T.F., Vese, L.A.: Active contours without edges. IEEE Trans. Image Process. 10, 266–277 (2001)
15. Bunce, S., Izzetoglu, M., Izzetoglu, K., Onaral, B., Pourrezaei, K.: Functional near-infrared spectroscopy: an emerging neuro-imaging modality. Eng. Med. Biol. 25, 54–62 (2006)
16. Sassaroli, A., Frederick, B., Tong, Y., Renshaw, P.F., Fantini, S.: Spatially weighted BOLD signal for comparison of functional magnetic resonance imaging and near-infrared imaging of the brain. Neuroimage 33, 505–514 (2006)
17. Dehaes, M., Kazemi, K., Pelegrini-Issac, M., Grebe, R., Benali, H., Wallois, F.: Quantitative effect of the neonatal fontanel on synthetic near infrared spectroscopy measurements. Hum. Brain Mapp. (2011) (to appear), doi:10.1002/hbm.21483
18. Holm, S.: A simple sequentially rejective multiple test procedure. Scand. J. Statist. 6, 65–70 (1979)

Hemodynamic-Informed Parcellation of fMRI Data in a Joint Detection Estimation Framework

L. Chaari[1], F. Forbes[1], T. Vincent[1], and P. Ciuciu[2]

[1] Mistis team, Inria Grenoble and LJK, France
[2] CEA/DSV/I²BM/Neurospin, LNAO, Gif-Sur-Yvette, France

Abstract. Identifying brain hemodynamics in event-related functional MRI (fMRI) data is a crucial issue to disentangle the vascular response from the neuronal activity in the BOLD signal. This question is usually addressed by estimating the so-called Hemodynamic Response Function (HRF). Voxelwise or region-/parcelwise inference schemes have been proposed to achieve this goal but so far all known contributions commit to pre-specified spatial supports for the hemodynamic territories by defining these supports either as individual voxels or a priori fixed brain parcels. In this paper, we introduce a Joint Parcellation-Detection-Estimation (JPDE) procedure that incorporates an adaptive parcel identification step based upon local hemodynamic properties. Efficient inference of both evoked activity, HRF shapes and *supports* is then achieved using variational approximations. Validation on synthetic and real fMRI data demonstrate the JPDE performance over standard detection estimation schemes and suggest it as a new brain exploration tool.

1 Introduction

Within-subject analysis in event-related BOLD fMRI mainly relies on (i) *detection* of evoked activity to localize which parts of the brain are activated by a given stimulus type, and on (ii) *estimation* of the dynamics of the brain response also known as the Hemodynamic Response Function (HRF). Most approaches to detect neural activity rely on a single *a priori* HRF model for the whole brain although there has been evidence that this response can vary between cortical regions and across subjects [8] and that an accurate HRF model may significantly improve detection performance. To capture this variability, robust HRF estimation is necessary which can be achieved only in voxels or regions that elicit an evoked response to a given stimulus [9]. So far, many works have addressed this issue either by considering linear or nonlinear HRF models [1,4,14], parametric, semi-parametric or non-parametric (*i.e.* FIR models) descriptions [6,16,7], and by performing univariate (voxelwise) [4,16], multivariate (regionwise) [10,13] or even multiscale, *i.e.* spatially adaptive inference [15]. However, to the best of our knowledge, all these existing works assume the spatial support of the HRFs, either defined at the voxel or region-level, to be pre-specified. The proposed methodology takes place in the Joint Detection-Estimation (JDE) framework introduced in [10] and extended in [13,3] to account for spatial correlation between

voxels. Standard JDE-based inference requires a pre-specified decomposition of the brain into functionally homogeneous parcels (groups of connected voxels) but with no guarantee of their optimality. These parcels should be small enough to guarantee the invariance of the HRF within each parcel but large enough to contain reliable information for its inference [12]. Here, we introduce the concept of *hemodynamic territory* as a set of parcels which share a common HRF pattern. To determine such sets, we incorporate an additional layer in the JDE hierarchy, namely an adaptive parcel identification step based upon local hemodynamic properties. In this novel Joint Parcellation-Detection-Estimation (JPDE) model (Section 2), for all the parcels of a given territory, HRFs are voxelwise but defined as local stochastic perturbations of the same HRF pattern. Then, hemodynamics estimation reduces to the identification of a limited number (say K) of such HRF patterns and parcel identification reformulates as a clustering problem where each voxel is assigned an HRF group among K. The HRF group assignment variables are governed by a hidden Markov Model to enforce spatial correlation, *i.e.* favor group assignments to vary smoothly. Finally, the overall scheme iteratively identifies hemodynamic territories as pairs of one HRF pattern and a set of parcels assigned to the corresponding HRF group.

The proposed approach thus makes the JDE framework fully adaptive and more flexible. It is based on a variational Expectation Maximization (EM) algorithm (Section 3) to derive estimates of the HRF patterns, the response amplitude, the corresponding labels (activating/non-activating voxels) and the HRF group labels. Results on artificial and real fMRI data demonstrate that the JPDE approach outperforms the standard JDE (see Section 4).

2   A Joint Parcellation-Detection-Estimation model

2.1   Observed and Missing Variables

We extend the parcel-based JDE model of [10,13] to a whole-brain one, with a set of voxels denoted by $\boldsymbol{\mathcal{P}}$, and recast it in a missing data framework. At voxel j, the fMRI time series $\boldsymbol{y}_j$ is measured at times $\{t_n, n = 1{:}N\}$, where $t_n = n\,TR$, N being the number of scans and TR the time of repetition. The number of different stimulus types or experimental conditions is M. At each voxel j, we assume a voxel dependent HRF $\boldsymbol{h}_j \in \mathbb{R}^{D+1}$ with $\boldsymbol{H} = \{\boldsymbol{h}_j, j \in \boldsymbol{\mathcal{P}}\}$ the set of all HRFs. Each $\boldsymbol{h}_j$ is associated with a HRF group among K. These groups or HRF classes are specified by a set of hidden labels $\boldsymbol{Z} = \{z_j, j \in \boldsymbol{\mathcal{P}}\}$ where $z_j \in \{1 : K\}$ and $z_j = k$ means that voxel j belongs to the k-th group. An estimation of $\boldsymbol{Z}$ corresponds then to a partition of the brain into K hemodynamic territories whose connected components define a parcellation. The link to the observed BOLD data is specified via the following forward model:

$$\forall j \in \boldsymbol{\mathcal{P}}, \quad \boldsymbol{y}_j = \sum_{m=1}^{M} a_j^m \boldsymbol{X}_m \boldsymbol{h}_j + \boldsymbol{P}\boldsymbol{\ell}_j + \boldsymbol{\varepsilon}_j, \tag{1}$$

where the binary matrix $\boldsymbol{X}_m = \{x_m^{n-d\Delta t}, n = 1 : N, d = 0 : D\}$ is of size $N \times (D+1)$ and provides information on the stimulus occurrences for the

m-th experimental condition, $\Delta t < TR$ being the sampling period of the unknown HRFs. The scalar a_j^m's are weights that model the transition between stimulations and the neuro-vascular response. They are generally referred to as Neural Response Levels (NRL). We denote by $\boldsymbol{A} = \{\boldsymbol{a}^m, m = 1 : M\}$ with $\boldsymbol{a}^m = \{a_j^m, j \in \mathcal{P}\}$ the response amplitudes, a_j^m being the amplitude at voxel j for condition m. Similarly to the HRF's, each NRL is assumed to be in one of I groups specified by activation class assignment variables $\boldsymbol{Q} = \{\boldsymbol{q}^m, m = 1 : M\}$ where $\boldsymbol{q}^m = \{q_j^m, j \in \mathcal{P}\}$ and q_j^m represents the *activation class* at voxel j for condition m. The number of classes considered here is $I = 2$ for activated ($i = 2$) and non-activated ($i = 1$) voxels. Finally, the rest of the signal is made of vector $\boldsymbol{P}\ell_j$, which corresponds to low frequency drifts with $\boldsymbol{P}$ a $N \times O$ matrix, $\ell_j \in \mathbb{R}^O$ a vector to be estimated and $\boldsymbol{L} = \{\ell_j, j \in \mathcal{P}\}$. Regarding the observation noise, the $\boldsymbol{\varepsilon}_j$'s are assumed to be independent with $\boldsymbol{\varepsilon}_j \sim \mathcal{N}(0, \boldsymbol{\Gamma}_j^{-1})$. The set of all unknown precision matrices is denoted by $\boldsymbol{\Gamma} = \{\boldsymbol{\Gamma}_j, j \in \mathcal{P}\}$.

2.2   Hierarchical Model of the Complete Data Distribution

With standard additional assumptions [10,13,3], the joint model distribution writes $p(\boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{H}, \boldsymbol{Q}, \boldsymbol{Z}) = p(\boldsymbol{Y} \,|\, \boldsymbol{A}, \boldsymbol{H}) \, p(\boldsymbol{A} \,|\, \boldsymbol{Q}) \, p(\boldsymbol{Q}) \, p(\boldsymbol{H} \,|\, \boldsymbol{Z}) \, p(\boldsymbol{Z})$.

Likelihood. Assuming spatial independence of the noise, the likelihood reads $p(\boldsymbol{Y} \,|\, \boldsymbol{A}, \boldsymbol{H}; \boldsymbol{L}, \boldsymbol{\Gamma}) \propto \prod_{j \in \mathcal{P}} \mathcal{N}\big(\boldsymbol{y}_j; \sum_{m=1}^{M} a_j^m \boldsymbol{X}_m \boldsymbol{h}_j + \boldsymbol{P}\ell_j, \boldsymbol{\Gamma}_j^{-1}\big)$. Various possibilities for the $\boldsymbol{\Gamma}_j$'s include standard white and autoregressive noise models [10].

Neuronal Response Levels. The NRLs are assumed to be statistically independent across conditions: $p(\boldsymbol{A}; \boldsymbol{\theta_a}) = \prod_{m=1}^{M} p(\boldsymbol{a}^m; \boldsymbol{\theta}_m)$ where $\boldsymbol{\theta_a} = \{\boldsymbol{\theta}_m, m = 1 : M\}$ and $\boldsymbol{\theta}_m$ gathers the parameters for the m-th condition. A mixture model is then adopted by using the allocation variables q_j^m to segregate non-activated voxels ($q_j^m = 1$) from activated ones ($q_j^m = 2$). For the m-th condition, and conditionally to the assignment variables $\boldsymbol{q}^m$, the NRLs are assumed to be independent: $p(\boldsymbol{a}^m \,|\, \boldsymbol{q}^m; \boldsymbol{\theta}_m) = \prod_{j \in \mathcal{P}} p(a_j^m \,|\, q_j^m; \boldsymbol{\theta}_m)$ with $p(a_j^m \,|\, q_j^m = i; \boldsymbol{\theta}_m) \sim \mathcal{N}(\mu_{mi}, v_{mi})$ and $\boldsymbol{\theta}_m = \{\mu_{mi}, v_{mi}, i = 1, 2\}$. We also denote $\boldsymbol{\mu} = \{\mu_{mi}, m = 1 : M, i = 1, 2\}$ and $\boldsymbol{v} = \{v_{mi}, m = 1 : M, i = 1, 2\}$. For non-activating voxels ($i = 1$) we set for all m, $\mu_{m1} = 0$. The other parameters are unknown and have to be estimated.

Activation Classes. We assume prior independence between the M experimental conditions regarding the activation class assignments: $p(\boldsymbol{Q}) = \prod_{m=1}^{M} p(\boldsymbol{q}^m; \beta_m)$. Also, the density $p(\boldsymbol{q}^m; \beta_m) \propto \exp\big(\beta_m U(\boldsymbol{q}^m)\big)$ defines a spatial Markov prior, namely an Ising model with interaction parameter β_m and energy function $U(\boldsymbol{q}^m) = \sum_{j \sim j'} \delta(q_j^m, q_{j'}^m)$ where $\forall (a, b) \in \mathbb{R}^2$, $\delta(a, b) = 1$ if $a = b$ and 0 otherwise. The notation $j \sim j'$ means that the summation is over all neighboring voxels (in a 6-connexity 3D neighborhood). The unknown parameters are denoted by $\boldsymbol{\beta} = \{\beta_m, m = 1 : M\}$.

HRF Groups. In order to promote parcellation connexity, we also introduce here a spatial Markov prior, namely a K-class Potts model with interaction parameter β_z: $p(\boldsymbol{Z}; \beta_z) \propto \exp\big(\beta_z U(\boldsymbol{Z})\big)$, where the global energy reads $U(\boldsymbol{Z}) = \sum_{j \sim j'} \delta(z_j, z_{j'})$, i.e. neighboring voxels tend to belong to the same HRF group.

HRF Patterns. In contrast to [10, 13, 3] where a unique HRF shape is considered for a whole parcel, the distribution of $\boldsymbol{h}_j$ is expressed, for each voxel j, conditionally to the HRF group variable z_j: $p(\boldsymbol{H} | \boldsymbol{Z}) = \prod_{j \in \mathcal{P}} p(\boldsymbol{h}_j | z_j)$ with $p(\boldsymbol{h}_j | z_j = k) \sim \mathcal{N}(\bar{\boldsymbol{h}}_k, \bar{\boldsymbol{\Sigma}}_k)$. Here, the mean vector $\bar{\boldsymbol{h}}_k$ can be seen as the HRF pattern for group k and $\bar{\boldsymbol{\Sigma}}_k = v_h I_D$ regulates the stochastic perturbations around $\bar{\boldsymbol{h}}_k$. In addition, smooth $\bar{\boldsymbol{h}}_k$'s are favored by controlling their second order derivatives: $\bar{\boldsymbol{h}}_k \sim \mathcal{N}(\boldsymbol{0}, \sigma_h^2 \boldsymbol{R})$ with $\boldsymbol{R} = (\Delta t)^4 \, (\boldsymbol{D}_2^{\mathrm{t}} \boldsymbol{D}_2)^{-1}$ where $\boldsymbol{D}_2$ is the second-order finite difference matrix and σ_h^2 is a parameter to be estimated or fixed. Moreover, $\bar{\boldsymbol{h}}_{k0} = \bar{\boldsymbol{h}}_{kD\Delta t} = 0$ as in [10, 13, 3]. The parameters are then denoted by $\boldsymbol{\Theta} = \big\{ \boldsymbol{\Gamma}, \boldsymbol{L}, \boldsymbol{\mu}, \boldsymbol{v}, \boldsymbol{\beta}, \beta_z, \sigma_h^2, (\bar{\boldsymbol{h}}_k, \bar{\boldsymbol{\Sigma}}_k)_{1 \leq k \leq K} \big\}$ and belong to a set $\underline{\boldsymbol{\Theta}}$.

3   Variational EM Estimation

We propose to use an EM framework to deal with the missing data $\boldsymbol{A} \in \mathcal{A}$, $\boldsymbol{H} \in \mathcal{H}$, $\boldsymbol{Q} \in \mathcal{Q}$, $\boldsymbol{Z} \in \mathcal{Z}$. We resort to an iterative variational EM procedure as in [3]. At each iteration (r), with $\boldsymbol{\Theta}^{(r-1)}$ denoting the current parameter values, the intractable posterior $p(\boldsymbol{A}, \boldsymbol{H}, \boldsymbol{Q}, \boldsymbol{Z} | \boldsymbol{Y}, \boldsymbol{\Theta}^{(r-1)})$ is approximated as a product of four pdfs, $\widetilde{p}_H^{(r)}$, $\widetilde{p}_A^{(r)}$, $\widetilde{p}_Q^{(r)}$ and $\widetilde{p}_Z^{(r)}$ respectively on $\mathcal{A}$, $\mathcal{H}$, $\mathcal{Q}$ and $\mathcal{Z}$. Our E-step becomes then an approximate E-step, which is decomposed into four sub-steps that consist of updating the four pdfs above in turn. Compared to [3], this implies adding an E-sub-step for the HRF group assignments ($\widetilde{p}_Z^{(r)}$ updating) and specifying its impact on the other E-sub-steps. The **E-Q** sub-step ($\widetilde{p}_Q^{(r)}$ updating) is not actually impacted by the HRF groups addition and can be found in [3]. The **E-A** sub-step ($\widetilde{p}_A^{(r)}$ updating) is also very close to the one involved in [3]: similar updating formulas are obtained by replacing the HRF of [3] by voxel dependent HRFs. We thus only detail the **E-H** and **E-Z** steps. At iteration (r), with current estimates $\tilde{p}_A^{(r-1)}$, $\tilde{p}_Z^{(r-1)}$ and $\boldsymbol{\Theta}^{(r-1)}$, we obtain:

$$\textbf{E-H:} \; \widetilde{p}_H^{(r)}(\boldsymbol{H}) \propto \exp\left(\mathrm{E}_{\tilde{p}_A^{(r-1)} \tilde{p}_Z^{(r-1)}}\left[\log p(\boldsymbol{H} | \boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{Z}; \boldsymbol{\Theta}^{(r-1)}) \right] \right) \tag{2}$$

$$\textbf{E-Z:} \; \widetilde{p}_Z^{(r)}(\boldsymbol{Z}) \propto \exp\left(\mathrm{E}_{\tilde{p}_H^{(r)}}\left[\log p(\boldsymbol{Z} | \boldsymbol{Y}, \boldsymbol{H}; \boldsymbol{\Theta}^{(r-1)}) \right] \right), \tag{3}$$

where $\mathrm{E}_{\tilde{p}}[.]$ denotes the expectation with respect to $\tilde{p}$.

It follows from standard algebra that $\tilde{p}_H^{(r)}$ and $\widetilde{p}_A^{(r-1)}$ are both Gaussian distributions: $\widetilde{p}_H^{(r)} = \prod_{j \in \mathcal{P}} \widetilde{p}_{H_j}^{(r)}$ and $\widetilde{p}_A^{(r-1)} = \prod_{j \in \mathcal{P}} \widetilde{p}_{A_j}^{(r-1)}$, where $\widetilde{p}_{H_j}^{(r)} \sim \mathcal{N}(\boldsymbol{m}_{H_j}^{(r)}, \boldsymbol{\Sigma}_{H_j}^{(r)})$ and $\tilde{p}_{A_j}^{(r-1)} \sim \mathcal{N}(\boldsymbol{m}_{A_j}^{(r-1)}, \boldsymbol{\Sigma}_{A_j}^{(r-1)})$. More specifically, we obtain:

• **E-H step:** Compute $\boldsymbol{\Sigma}_{H_j}^{(r)} = (\boldsymbol{V}_1 + \boldsymbol{V}_2)^{-1}$ and $\boldsymbol{m}_{H_j}^{(r)} = \boldsymbol{\Sigma}_{H_j}^{(r)}(\boldsymbol{m}_1 + \boldsymbol{m}_2)$, where $\boldsymbol{V}_1 = \displaystyle\sum_{m,m'} \boldsymbol{\Sigma}_{A_j(m,m')}^{(r-1)} \boldsymbol{X}_m^{\mathrm{t}} \boldsymbol{\Gamma}_j^{(r-1)} \boldsymbol{X}_{m'} + \widetilde{\boldsymbol{S}}_j^{\mathrm{t}} \boldsymbol{\Gamma}_j^{(r-1)} \widetilde{\boldsymbol{S}}_j$, $\boldsymbol{V}_2 = \displaystyle\sum_{k=1}^{K} \widetilde{p}_{Z_j}(k)^{(r-1)} \bar{\boldsymbol{\Sigma}}_k^{(r-1)-1}$, $\boldsymbol{m}_1 = \widetilde{\boldsymbol{S}}_j^{\mathrm{t}} \boldsymbol{\Gamma}_j^{(r-1)}(\boldsymbol{y}_j - \boldsymbol{P}\boldsymbol{\ell}_j^{(r-1)})$ and $\boldsymbol{m}_2 = \sum_{k=1}^{K} \bar{\boldsymbol{\Sigma}}_k^{(r-1)-1} \widetilde{p}_{Z_j}(k)^{(r-1)} \bar{\boldsymbol{h}}_k^{(r-1)}$. Above, $\widetilde{\boldsymbol{S}}_j = \sum_{m=1}^{M} m_{A_j^m}^{(r-1)} \boldsymbol{X}_m$ and $m_{A_j^m}^{(r-1)}$, $\boldsymbol{\Sigma}_{A_j(m,m')}^{(r-1)}$ denote respectively the m and (m,m') entries of $\boldsymbol{m}_{A_j}^{(r-1)}$ and $\boldsymbol{\Sigma}_{A_j}^{(r-1)}$.

• **E-Z step:** Akin to [3], we resort to a mean field approximation, $\widetilde{p}_Z^{(r)}(\boldsymbol{Z}) = \prod_{j \in \mathcal{P}} \widetilde{p}_{Z_j}^{(r)}(z_j)$ where $\widetilde{p}_{Z_j}^{(r)}(k) \propto \mathcal{N}(\boldsymbol{m}_{H_j}^{(r)}; \bar{h}_k^{(r-1)}, \bar{\boldsymbol{\Sigma}}_k^{(r-1)}) \exp\{\frac{-1}{2}\mathrm{trace}(\boldsymbol{\Sigma}_{H_j}^{(r)} \bar{\boldsymbol{\Sigma}}_k^{-1}) + \beta_z \sum_{l \sim j} \delta(k, \widetilde{z}_l)\}$, where $\{\widetilde{z}_j, j \in \mathcal{P}\}$ is a particular configuration of $\boldsymbol{Z}$ updated according to a specific scheme [2] and $\sim j$ denotes voxels neighboring j.

• **M step:** The maximization step can also be divided into five sub-steps (two additional ones compared to [3]) involving separately $(\boldsymbol{\mu}, \boldsymbol{v})$, $\boldsymbol{\beta}$, β_z, $(\mathbf{L}, \boldsymbol{\Gamma})$ and $(\bar{\boldsymbol{h}}_k, \bar{\boldsymbol{\Sigma}}_k)_{1 \leq k \leq K}$. For the $(\boldsymbol{\mu}, \boldsymbol{v})$ and $(\bar{\boldsymbol{h}}_k, \bar{\boldsymbol{\Sigma}}_k)_{1 \leq k \leq K}$ sub-steps, closed forms can be analytically derived for the updates. Numerical procedures are required for the other sub-steps. See [3] for details.

4   Validation

Artificial Datasets. Experiments have been carried out on artificial fMRI data generated according to Eq. (1). We simulated a random mixed sequence of indexes coding for $M = 2$ different stimuli composed of 30 trials each. The resulting ternary sequence was then multiplied by stimulus-dependent and space-varying NRLs, which were drawn from the prior distribution $p(\boldsymbol{A}; \boldsymbol{\theta}_a)$. To this end, 2D slices composed of 20 x 20 binary labels $\boldsymbol{Q}^m$ (activating and non-activating voxels) were constructed for each stimulus type m (see Fig. 1[Left]). Given these labels, the NRLs were simulated as follows, for $m = 1, 2$: $a_j^m \mid q_j^m = 1 \sim \mathcal{N}(0, 0.5)$ and $a_j^m \mid q_j^m = 2 \sim \mathcal{N}(3.2, 0.5)$ (see Fig. 2[Left]). As regards HRFs, three groups $(K = 3)$ were considered and spatially organized in three parcels of similar size (labels $\boldsymbol{Z}$) as shown in Fig. 1[Top-right]. Within each parcel, all voxels share the same HRF prior parameters $(\bar{\boldsymbol{h}}_k, \bar{\boldsymbol{\Sigma}}_k)$. The mean HRF shapes $(\bar{\boldsymbol{h}}_k)_{k=1:K}$ are depicted in Fig. 3 and show strong fluctuations across parcels. Diagonal prior covariance matrices $(\bar{\boldsymbol{\Sigma}}_k)_{k=1:K}$ were considered to draw voxel-specific HRFs according to $p(\boldsymbol{h}_j | z_j = k)$.

As regards parcellation, Fig. 1[Top-right] shows the ability of JPDE to recover the spatial support of hemodynamic territories with high accuracy (1% of misclassified voxels and a DICE index of 0.993) from an imperfect initialization (Fig. 1[Bottom-right]). The HRF variability does not seem to affect the activation maps which are equally well estimated in the JPDE and JDE cases (Fig. 1[Left]). However, a clear difference is seen on the estimated HRFs, which are depicted in Fig. 3 together with the ground truth: the three parcel-specific HRF estimates using JPDE are plotted as well as the single JDE-based HRF time course obtained by merging all parcels. The JPDE estimation is accurate

for all parcels although the parcels cover different proportions of activation areas (*i.e.* useful signal). In contrast, JDE provides an intermediate HRF shape which lies between those of the three parcels. This explains the observed differences between the two models in terms of estimated NRL dynamics and points out the JDE sensitivity to the choice of the *a priori* parcellation. When imperfect,
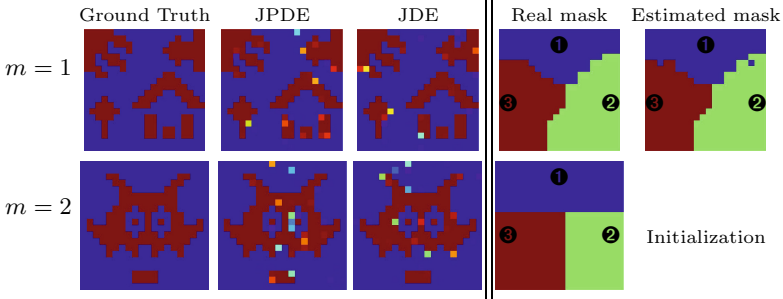

Fig. 1. Left: reference activation labels and Posterior Probability Maps (PPM) for JPDE and JDE (a single parcel is assumed for JDE); **Right:** reference, estimated and initial parcellation masks
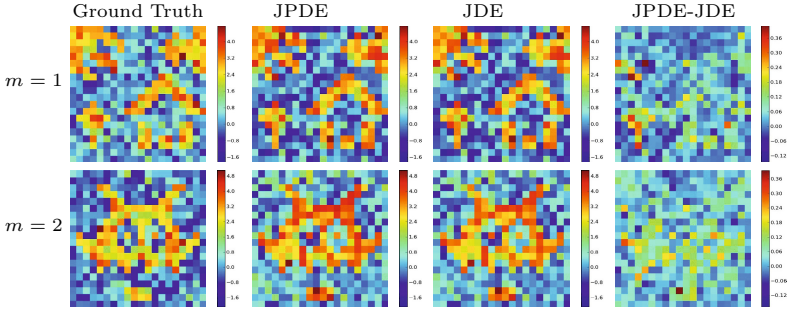

Fig. 2. Reference and estimated NRLs using JPDE (3 parcels) and JDE (1 parcel)
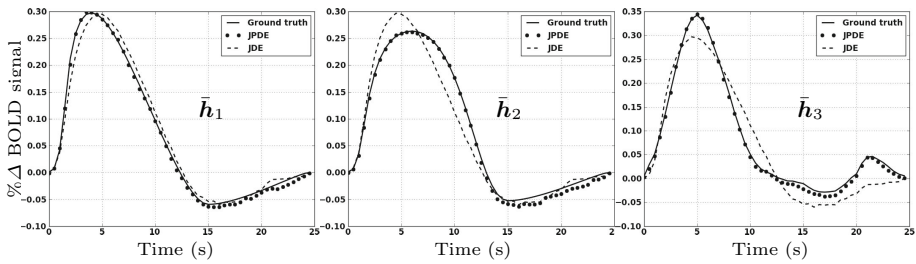

Fig. 3. Reference and estimated HRF patterns ($\bar{h}_k$) for each parcel using JPDE and JDE

JDE is forced to miss-fit the real HRF shape, and therefore activation dynamics. In the same context, JPDE is able to automatically refine an initial parcellation and provide reliable detection and estimation results.

Interestingly, the NRL differences in Fig. 2 (see the JPDE-JDE plots in Fig. 2[Right]) show that NRL estimates with JPDE have higher pic values, which means that JPDE allows retrieving stronger activation dynamics closer to the ground truth. The most significant NRL differences lie in parcels 2 and 3 where the JDE HRF estimate differs the most from the ground truth. In terms of Mean Square Error (MSE), reported values confirm the performance of JPDE over JDE: $MSE_{\text{JDE}}^{m=1} = 0.0182$ *vs* $MSE_{\text{JPDE}}^{m=1} = 0.0107$ and $MSE_{\text{JDE}}^{m=2} = 0.0183$ *vs* $MSE_{\text{JPDE}}^{m=2} = 0.0141$.

Real Data. fMRI data were recorded at 3 T (Siemens Trio) using a gradient-echo EPI sequence (TE=30ms/TR=2.4s/thickness=3mm/FOV=192mm^2) during a *Localizer* experiment [11] with a fast event-related paradigm. The paradigm involved sixty auditory (Aud.), visual (Vis.) and motor stimuli, defined in ten experimental conditions (Aud./Vis. sentences, Aud./Vis. calculations, left/right Aud. and Vis. clicks, horizontal and vertical checkerboards). For the considered dataset, the acquisition consisted of a single session of $N = 128$ scans, yielding 3-D volumes with a spatial resolution of $2 \times 2 \times 3\text{mm}^3$. In this experiment, we focus on the `Auditory` condition which is supposed to reveal activations in the temporal lobes. The initial parcellation used (from [12]) and the JPDE estimated one are shown in Fig. 4[Top-middle]. It appears that JPDE groups a number of initial parcels as they turn out to have similar hemodynamic properties, which suggests that the initial parcellation may be unnecessarily too fine. JPDE retrieves respectively three and two different parcels in the left and right temporal regions of interest (ROI). However, JPDE HRF estimates (Fig. 4[Top-right]) show very close shapes for parcels 1 and 3 for the left ROI, which explains
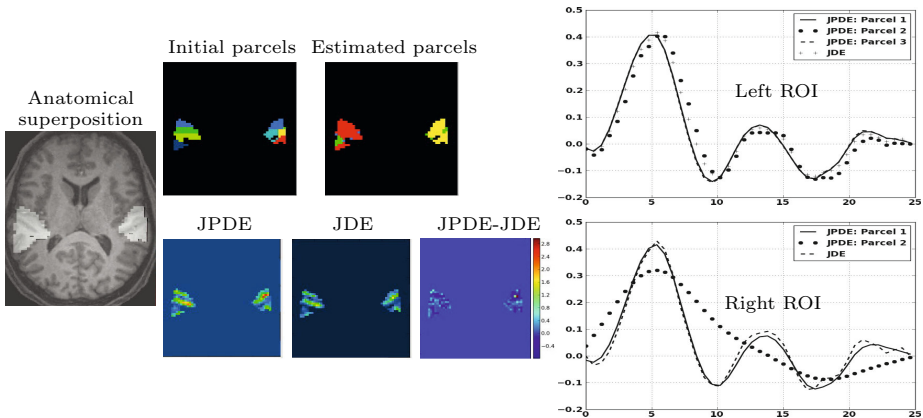

Fig. 4. Left: ROI definition; **Middle:** Initial and estimated parcels (top), NRL estimates with JDE and JPDE and difference image (bottom); **Right:** HRF estimates for the estimated parcels

the reduced size of the third parcel. As regards activation levels, Fig. 4[Bottom-middle] shows the estimated NRLs using JPDE and JDE. The difference image in Fig. 4 confirms the ability of JPDE to retrieve stronger activations w.r.t. to JDE.

5   Conclusion

We proposed a JPDE framework that provides an automatic parcellation of the brain into homogeneous hemodynamic territories. The quality and reliability of such a parcellation is at the core of robust neural activity detection and brain hemodynamics estimation. By enabling a fully adaptive data-dependent iden-tification of the parcels, the JPDE framework greatly extends the possibilities of detection-estimation approaches. The gain in removing the commitment to a priori fixed territories has been confirmed in preliminary experiments that showed that the JPDE achieved better results than the standard JDE using a fixed parcellation. An important remaining question raised by this new frame-work is related to the issue of choosing the right number of HRF groups at best *i.e.* in a sparse manner so as to capture the spatial variability in hemodynamic territories while enabling the reproducibility of parcel identification across fMRI datasets. This question should be the most critical to validate our approach but also the most interesting to neuroscientists in case of success. For this specific point, we shall investigate variational approximations of standard information criteria [5] such as the Bayesian Information Criterion.

References

1. Boynton, G.M., Engel, S.A., Glover, G.H., Heeger, D.J.: Linear systems analysis of functional magnetic resonance imaging in human V1. J. Neurosci. 16, 4207–4221 (1996)
2. Celeux, G., Forbes, F., Peyrard, N.: EM procedures using mean field-like approx-imations for Markov model-based image segmentation. Patt. Rec. 36, 131–144 (2003)
3. Chaari, L., Forbes, F., Vincent, T., Dojat, M., Ciuciu, P.: Variational Solution to the Joint Detection Estimation of Brain Activity in fMRI. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part II. LNCS, vol. 6892, pp. 260–268. Springer, Heidelberg (2011)
4. Ciuciu, P., Poline, J., Marrelec, G., Idier, J., Pallier, C., Benali, H.: Unsupervised robust non-parametric estimation of the hemodynamic response function for any fMRI experiment. IEEE Trans. Med. Imag. 22(10), 1235–1251 (2003)
5. Forbes, F., Peyrard, N.: Hidden Markov Random Field model selection crite-ria based on mean field-like approximations. IEEE Trans. Patt. Anal. Mach. In-tell. 25(9), 1089–1101 (2003)
6. Glover, G.H.: Deconvolution of impulse response in event-related BOLD fMRI. Neuroim. 9, 416–429 (1999)
7. Goutte, C., Nielsen, F., Hansen, L.K.: Modeling the haemodynamic response in fMRI using smooth FIR filters. IEEE Trans. Med. Imag. 19(12), 1188–1201 (2000)

8. Handwerker, D.A., Ollinger, J.M., Mark, D.: Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. Neuroim. 21, 1639–1651 (2004)
9. Kershaw, J., Ardekani, B.A., Kanno, I.: Application of Bayesian inference to fMRI data analysis. IEEE Trans. Med. Imag. 18(12), 1138–1152 (1999)
10. Makni, S., Idier, J., Vincent, T., Thirion, B., Dehaene-Lambertz, G., Ciuciu, P.: A fully Bayesian approach to the parcel-based detection-estimation of brain activity in fMRI. Neuroim. 41(3), 941–969 (2008)
11. Pinel, P., Thirion, B., Mériaux, S., Jobert, A., Serres, J., Le Bihan, D., Poline, J.B., Dehaene, S.: Fast reproducible identification and large-scale databasing of individual functional cognitive networks. BMC Neurosci. 8(1), 91 (2007)
12. Thirion, B., Flandin, G., Pinel, P., Roche, A., Ciuciu, P., Poline, J.B.: Dealing with the shortcomings of spatial normalization: Multi-subject parcellation of fMRI datasets. Hum. Brain Mapp. 27(8), 678–693 (2006)
13. Vincent, T., Risser, L., Ciuciu, P.: Spatially adaptive mixture modeling for analysis of within-subject fMRI time series. IEEE Trans. Med. Imag. 29, 1059–1074 (2010)
14. Wager, T.D., Vazquez, A., Hernandez, L., Noll, D.C.: Accounting for nonlinear BOLD effects in fMRI: parameter estimates and a model for prediction in rapid event-related studies. Neuroim. 25(1), 206–218 (2005)
15. Wang, J., Zhu, H., Fan, J., Giovanello, K., Lin, W.: Adaptively and Spatially Estimating the Hemodynamic Response Functions in fMRI. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part II. LNCS, vol. 6892, pp. 269–276. Springer, Heidelberg (2011)
16. Woolrich, M., Jenkinson, M., Brady, J., Smith, S.: Fully Bayesian spatio-temporal modelling of fMRI data. IEEE Trans. Med. Imag. 23(2), 213–231 (2004)

Group Analysis of Resting-State fMRI by Hierarchical Markov Random Fields

Wei Liu, Suyash P. Awate, and P. Thomas Fletcher

Scientific Computing and Imaging Institute, University of Utah, USA
weiliu@sci.utah.edu

Abstract. Identifying functional networks from resting-state functional MRI is a challenging task, especially for multiple subjects. Most current studies estimate the networks in a sequential approach, i.e., they identify each individual subject's network independently to other subjects, and then estimate the group network from the subjects networks. This one-way flow of information prevents one subject's network estimation benefiting from other subjects. We propose a hierarchical Markov Random Field model, which takes into account both the within-subject spatial coherence and between-subject consistency of the network label map. Both population and subject network maps are estimated simultaneously using a Gibbs sampling approach in a Monte Carlo Expectation Maximization framework. We compare our approach to two alternative groupwise fMRI clustering methods, based on K-means and Normalized Cuts, using both synthetic and real fMRI data. We show that our method is able to estimate more consistent subject label maps, as well as a stable group label map.

1 Introduction

Resting-state functional MRI (rs-fMRI) is widely used for detecting the intrinsic functional networks of the human brain. The availability of large rs-fMRI databases opens the door for systematic group studies of functional connectivity. While the inherently high level of noise in fMRI makes functional network estimation difficult at the individual level, combining many subjects' data together and jointly estimating the common functional networks is more robust. However, this approach does not produce estimates of individual functional connectivity. Such individual estimates are an important step in understanding functional networks not just on average, but also how these networks vary across individuals.

The most common approaches for functional network identification are Independent Component Analysis (ICA) and its variants [2], which identify the statistically independent functional networks without *a priori* knowledge of the regions of interest. The more recently proposed clustering-based methods [1,8] partition the brain into disjoint spatial clusters, or label maps, representing the functional networks. Group ICA [2] is a generalization of ICA to multiple subjects, in which all subjects are assumed to share a common spatial component map but have distinct time courses. The time courses from all subjects

are concatenated temporally, followed by a single ICA. Although the subject component maps are obtained by a back-reconstruction procedure, there is no explicit statistical modeling of the variability between the group and subject component maps. Ng et. al [6] use group replicator dynamics (RD) to detect subject's sparse component maps, with group information integrated into each subject's RD process. In clustering-based methods, the subjects clusterings are usually averaged to obtain a group affinity matrix and are followed by a second level clustering on the group similarity matrix [1,8]. Because the group level clustering is conducted after subject level clustering, the clustering of one subject is unaware of the information from other subjects, as well as the group clustering.

In this paper we propose a Bayesian hierarchical model to identify the functional networks from rs-fMRI that includes both subject and population levels. We assume a group network label map that acts as a prior to the label maps for all subjects in the population. This Bayesian perspective provides a natural regularization of the estimation problem of a single subject using information from the entire population. The variability between the subjects and group are taken into account through the conditional distributions between group and subjects. The within-subject spatial coherence is modeled by a Markov Random Field (MRF). Both the group clustering and subject clusterings are estimated simultaneously with a Monte Carlo Expectation Maximization (MCEM) algorithm. The model is data-driven in that all parameters, regularized by two given hyperparameters, are estimated from the data, and the only parameter that must be specified is the number of networks.

Markov Random Fields have previously been used in fMRI analysis to model spatial context information [3,4]. However, to our knowledge, ours is the first hierarchical MRF applied to fMRI for modeling both group and individual networks. The model of Ng et al. [5] combines all subjects into a single MRF and bypasses the need for one-to-one voxel correspondence across subjects, but the edges are added directly between subjects without a group layer. In our model, a group layer network map is explicitly defined, and the consistency between subjects is encoded through adding edges between group and subjects labels. Our method differs from other clustering methods [1,8] in that their methods identify the subject's functional network patterns independently, without any knowledge of other subjects or group population. Instead, our method estimates both levels of network patterns simultaneously. The proposed approach can be seen as a counterpart on the clustering branch of the multi-subject dictionary learning algorithm [9], which also has a hierarchical model and a spatially smoothed sparsity prior on the group component map.

2   Hierarchical Model for Functional Networks

We define each subject's network label map as a Markov Random Field (MRF) with statistical dependency between spatially adjacent voxels. These connections act as a prior model favoring spatial coherence of functional regions. An additional group label map is defined on top of all subject label maps. The group

label map has the same Markov structure as the individuals, again to encourage spatial coherence of the functional regions in the group level. In addition, each voxel in the group is connected to the corresponding voxel of each subject. These connections model the relationship between the group and the individuals. Hence, all voxels of subjects and group label map are jointly connected into a single MRF. See Figure 1 for an illustration. More specifically, define a graph $\mathcal{A} = (\mathcal{S}, \mathcal{E})$, and the set of node $\mathcal{S} = (\mathcal{G}, \mathcal{H})$. $\mathcal{G}$ is the set of voxels in the group label map, and $\mathcal{H} = (\mathcal{H}_1, \ldots, \mathcal{H}_J)$ includes voxels for all of the J subjects' label maps. An edge $(s, t) \in \mathcal{E}$ is defined if 1) $s \in \mathcal{H}^j, t \in \mathcal{G}$ and s, t are at the same voxel location, or 2) if $s, t \in \mathcal{G}$, and s, t are spatial neighbors, or 3) $s, t \in \mathcal{H}^j$, and s, t are spatial neighbors. On each node $s \in \mathcal{S}$, a random variable $y_s \in \mathcal{L} = \{1, \cdots, L\}$ is defined to represent the functional network labels.

MRF Prior: Our MRF prior on the hierarchical model is essentially a Potts model with different weights for the within-subject connections and the connections between the group and individuals. Because of the equivalence of MRFs and Gibbs fields, we define our prior as $p(Y) = \frac{1}{Z} \exp\{-U(Y)\}$, where the energy function $U(Y)$ is given by

$$U(Y) = \sum_{s,r \in \mathcal{G}} \beta \psi(y_s, y_r) + \sum_{j=1}^{J} \left(\sum_{s \in \mathcal{G}, \tilde{s} \in \mathcal{H}^j} \alpha \psi(y_s, y_{\tilde{s}}) + \sum_{s,r \in \mathcal{H}^j} \beta \psi(y_s, y_r) \right).$$

Here ψ is a binary function that is zero when the two inputs are equal and one otherwise, and α and β are parameters determining the strength of the connections. This regularization encodes two physiologically meaningful *a priori* assumptions on the functional networks under investigation: 1) The networks are spatially coherent within single subject. This is modeled by the β term. 2) The networks are similar between subjects, and therefore between the group and subjects. This is modeled by the α term.

Likelihood Model: In the generative model, for any individual subject, the observed time course at each voxel is assumed to be generated from a distribution conditioned on the network label at that voxel. In fMRI analysis the time series at each voxel is usually normalized to be zero mean and unit norm, so the analysis is robust to shifts or scalings of the data. This results in the data being projected onto a high-dimensional unit sphere. After normalization, the sample correlation between two time series is equal to their inner product.

We use the notation $X = \{(x_1, \ldots, x_N) \mid x_s \in S^{p-1}\}$ to denote the set of *normalized* time series in p-sphere. Given Y, the random vectors x_s are conditional independent, hence $\log p(X|Y) = \sum_{s \in \mathcal{H}} \log p(x_s|y_s)$. The likelihood function $p(x_s|y_s)$ is naturally modeled by a von Mises-Fisher (vMF) distribution

$$f(x_s|y_s = l; \mu_l, \kappa_l) = C_p(\kappa_l) \exp\left(\kappa_l \mu_l^T x_s\right), \quad x_s \in S^{p-1}, \quad l \in \mathcal{L}, \qquad (1)$$

where for the cluster labeled l, μ_l is the mean direction, $\kappa_l \geq 0$ is the *concentration parameter*, and C_p is the normalization constant. The larger the κ_l, the greater the density concentrated around the mean direction.

3   Bayesian Inference

We solve the inference problem in a *maximum a posteriori* (MAP) framework. That is, given the observed time course data X, we estimate the posterior mode of $p(Y|X)$. This consists of the following components.

Parameter Estimation: In this data-driven model, we propose to estimate the parameters $\theta = \{\alpha, \beta, \kappa, \mu\}$ from the data using an Expectation Maximization (EM) algorithm. However, the high-dimensionality and dependency between spatially adjacent voxels in MRF make it infeasible to obtain a closed form solution of the expectation of $[\log p(X, Y)]$ with respect to $p(Y|X)$. Here we propose to approximate the expectation using Monte Carlo EM (MCEM), in which a sample, $(Y^1, \cdots, Y^M)$, generated from density $p(Y|X)$ is used to approximate the expected value by the empirical average $\frac{1}{M} \sum_{m=1}^{M} \log p(X, Y^m)$.

Gibbs Sampling: Gibbs sampling converts a multivariate sampling problem into a consecutive univariate sampling, hence is well adapted to draw the Monte Carlo samples from $p(Y|X)$. In our hierarchical structure, the sampling procedure is also done in a hierarchical way. At the image level, a sample of the group label map, $Y_{\mathcal{G}}^m$, is drawn given the previous subject label map, $Y_{\mathcal{H}_j}^{m-1}$. Next, a sample for each subject map, $Y_{\mathcal{H}_j}^m$, is generated given the previous group label map, $Y_{\mathcal{G}}^{m-1}$. At the voxel level, we can draw samples of the label y_s given the rest of nodes fixed, and update $y_s, \forall s \in S$. The conditional probability used to generate samples at the group and subject voxels are given as

$$p(y_s|y_{-s}, X) \propto \frac{1}{Z_s} \exp\{-U(y_s|y_{-s})\} \cdot p(x_s|y_s) = \frac{1}{Z_s} \exp\{-U_p(y_s|y_{\mathcal{N}_s}, x_s)\}$$

$$U_p = \alpha \sum_{j=1}^{J} \psi(y_s, y_{\tilde{s}}^j) + \beta \sum_{r \in \mathcal{N}s} \psi(y_s, y_r), \quad \forall s \in \mathcal{G}, \tag{2}$$

$$U_p = \alpha \psi(y_s, y_{\tilde{s}}) + \beta \sum_{r \in \mathcal{N}s} \psi(y_s, y_r) - \kappa_l \mu_l^\top x_s - \log C_p, \forall s \in \mathcal{H}_j, \tag{3}$$

where $-s$ is the set of all nodes excluding s, Z_s the normalization constant, U_p is the posterior energy, and $\mathcal{N}_s$ is the set of neighbor's of s. In our model we use 6-neighbor system in a 3D volume image. $y_{\tilde{s}}^j$ in (2) is the label of subject j's voxel with the same spatial location with s, and $y_{\tilde{s}}$ in (3) is the label of group's voxel with the same spatial location with s. Because of the dependency on previous samples, the sequence of samples will be a Markov Chain, hence our method falls into Markov Chian Monte Carlo (MCMC) sampling. After a sufficient burn-in period, a series of samples $Y^m, m = 1 \cdots M$ is saved for approximating the expectation $\mathbb{E}[\log p(X, Y)]$.

Pseudo Likelihood: To evaluate $\log p(X, Y^m; \theta) = \log p(Y^m; \theta) + \log p(X|Y^m; \theta)$ as a function of θ, we face the difficulty of evaluating the partition function Z in $p(Y^m)$. In practice the Gibbs field is approximated by pseudo-likelihood, which is defined as the product of the conditional distribution $p(y_s|y_{-s}), \forall s \in \mathcal{S}$. Therefore the energy function can be written as

$$U(Y) \approx \sum_{s \in \mathcal{S}} U(y_s | y_{-s}) = \sum_{s \in \mathcal{G}} \left(\alpha \sum_{j \in J} \psi(y_s, y_{\tilde{s}}^j) + \beta \sum_{r \in \mathcal{N}s} \psi(y_s, y_r) \right)$$
$$+ \sum_{j=1}^{J} \sum_{s \in \mathcal{H}^j} \left(\alpha \psi(y_s, y_{\tilde{s}}) + \beta \sum_{r \in \mathcal{N}_s} \psi(y_s, y_r) \right),$$

where $y_{\tilde{s}}^j$ and $y_{\tilde{s}}$ has the same definition with (2) and (3).

Hierarchical MRF Algorithm Using MCEM: With all the preparation above, parameter estimation can be done by maximizing $\frac{1}{M} \sum_{m=1}^{M} \log p(X, Y^m)$. The α and β in the MRF prior can be optimized by maximizing $\frac{1}{M} \sum_{m=1}^{M} \log p(Y^m)$ with a Newton-Raphson method. We assume a Gaussian prior distribution on α with hyper-parameters μ_α and σ_α, which is given manually and does not have significant impact on the model. In order for MCMC sampling to converge quickly to the posterior, we need a reasonably good initial network label map. Here the K-means clustering on a concatenated group dataset is used for the initial maps of both the group and subjects. After the EM parameter estimation iterations are done, an Iterated Conditional Modes (ICM) on the current sample map gives the final label maps. Putting this all together, the groupmrf algorithm to estimate the group and individual label maps is given in Algorithm 1.

Algorithm 1. Monte Carlo EM for group MRF

Data: Normalized fMRI, initial group label map
Result: Group label map $Y_{\mathcal{G}}$, subjects label map $Y_{\mathcal{H}}$, parameters $\{\alpha, \beta, \mu, \sigma\}$
while $\mathbb{E}[\log p(X, Y)]$ *not converge* **do**
  repeat
    foreach $s \in \mathcal{G}$ **do** Draw consecutive samples of y_s from $p(y_s | y_{\mathcal{N}_s}, X; \theta)$
    using (2) ;
    foreach $j = 1 \ldots J$ **do**
      foreach $s \in \mathcal{H}_j$ **do** Draw consecutive samples of y_s from
      $p(y_s | y_{\mathcal{N}_s}, X; \theta)$ using (3) ;
    Save sample Y^m after B burn-ins;
  until $B + M$ *times*;
  foreach $l = 1 \cdots L$ **do**
    Estimate $\{\mu_l, \kappa_l\}$ by maximizing $\frac{1}{M} \sum_{m=1}^{M} \log p(X | Y^m)$;
  Estimate $\{\alpha, \beta\}$ by maximizing $\frac{1}{M} \sum_{m=1}^{M} \log p(Y^m)$;
Run ICM on current samples to estimate final label maps.

---

4   Results and Conclusion

Three methods are compared in both synthetic data and *in vivo* data test. The first method is K-Means [1] applied on each subject's fMRI data, as well as on a group dataset constructed by concatenating all subjects time courses. To

alleviate the dependency on initial cluster centers, we run K-Means 20 times with different initial cluster centers generated by a K-means++ algorithm. The second method is a Normalized-Cuts algorithm (N-Cuts), following Van den Heuvel, et al. [8], which is applied in two stages. First N-Cuts is run on each subject's affinity matrix, as computed by pairwise correlation between time courses. Second, N-Cuts is applied on a group affinity matrix, which is computed by summing up all of the subjects' segmentation matrices. We use the Ncutclustering_9 toolbox [7], a newer version of the one used in [8]. The third method is our groupmrf approach applied on all subjects' fMRI data. The preprocessing are same for all three methods except that groupmrf use image data without spatial smoothing, while the other two use data smoothed by a standard 6mm Gaussian filter.
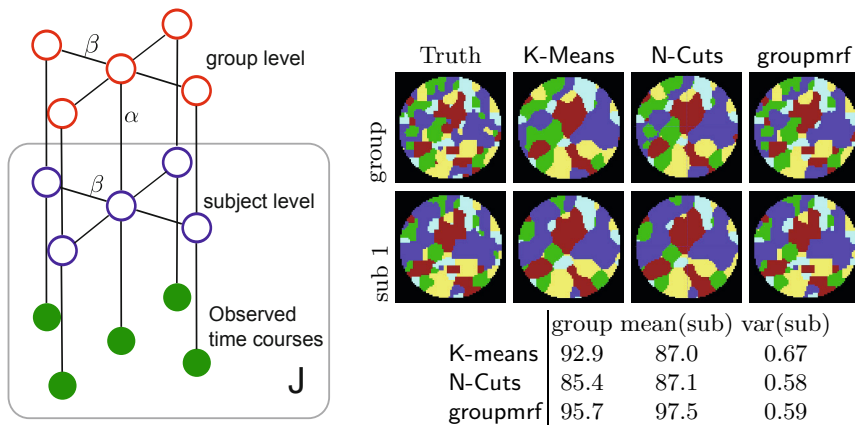

	group	mean(sub)	var(sub)
K-means	92.9	87.0	0.67
N-Cuts	85.4	87.1	0.58
groupmrf	95.7	97.5	0.59

Fig. 1. Left: Hierarchical MRF depicted by undirected graph. The J subjects are compactly represented by a box with label J. Right: clustering of K-means and N-Cuts on synthetic time series with spatial smoothing, and groupmrf without smoothing. Top is group label map and bottom is one of subjects label map. The table gives the rand index accuracy between estimated label map and ground truth image. The rand index of all subjects are summarized by a mean and variance value.

Synthetic Example: We simulate synthetic time course on each voxel of 16 subjects by first sampling from MRF with $\alpha = 0.4$ and $\beta = 2.0$ and get both group and subjects network label map. The time course signals at each voxel are generated by adding Gaussian white noise of $\sigma^2 = 40$ on each cluster's mean time course, which is synthesized from an auto-regressive process of $x_t = \varphi x_{t-1} + \varepsilon$ with $\varphi = 0.7$ and noise variance $\sigma_\varepsilon = 1$. The sample correlation between the mean time series is in the range of $(-0.15, 0.3)$. The rand index value on right side of Figure 1 shows that groupmrf algorithm is able to detect both group and subjects label map more accurately than the K-Means and N-Cuts method. The synthetic images shows that despite the different assumption of K-Means and N-cuts on the data, our algorithm is able to estimate subject label maps with more spatial and inter-subject coherence than the other two methods.
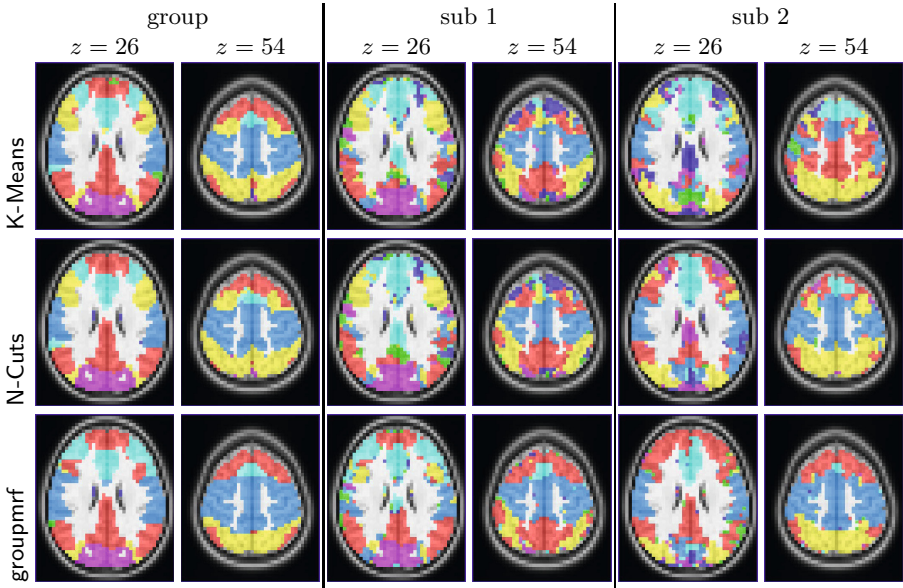
Fig. 2. Functional networks estimated by 3 methods shown in separate rows. groupmrf has more consistent estimation of the DMN (red) and motor network (blue) among two example subjects of 66 total used.

In Vivo Data: We tested our method on the ADHD-200 dataset in the 1000 Functional Connectomes Project. A total of 66 healthy control adolescent subjects were chosen from the same site (University of Pittsburgh). BOLD EPI images (TR = 1.5 s, TE = 29 ms, 29 slices at 4 mm slice thickness, 64 x 64 matrix, 196 volumes) were acquired on a Siemens 3 Tesla Trio scanner. The fMRI volumes were motion corrected, slice timing corrected, registered to NIHPD object 1 atlas, bandpass filtered to 0.01 to 0.1 Hz, regressed out nuisance variables including white matter, CSF mean time courses and six motion parameters, and at last filtered by a 8 mm Gaussian filter for spatial smoothness.

Figure 2 shows the functional networks computed from the three methods. As in the synthetic data experiment, all 66 subjects' time series were concatenated into a single group dataset. K-means and N-Cuts were applied on the spatially-smoothed, concatenated group dataset, as well as on each subject fMRI. Our groupmrf was applied on all subjects data without any spatial smoothing and with the initial parameter values $\alpha = 0.7, \beta = 1.0$. Following [8,1], the number of clusters are set to 7. It can be seen in Figure 2 that our algorithm is able to detect the major functional networks even for individual subjects, while K-means and N-Cuts miss some components in the Default Mode Network (DMN) for certain subjects, due to the high noise level of single subject data.

The real strength of Bayesian statistics lies in the probabilistic explanation of the results. The last experiment in Figure 3 shows the posterior probability maps of DMN and attention network of two subjects. The maps are approximated
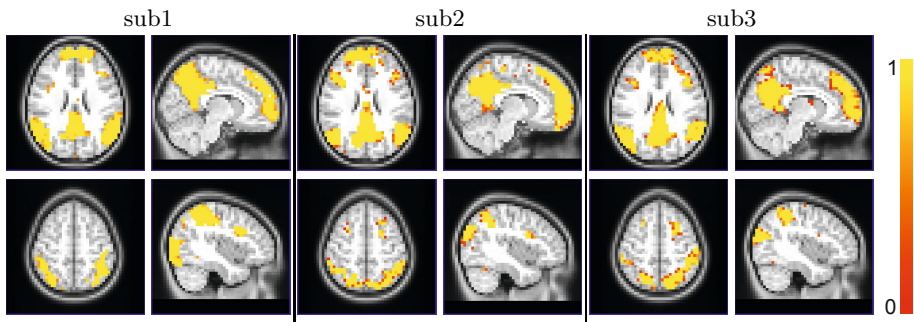
Fig. 3. Posterior probability maps of DMN and attention network for 3 example subjects out of the 66 total used. Top row: DMN, $x = -8, z = 26$. Bottom row: attention network. $x = 40, z = 54$.

by averaging the Monte-Carlo samples from the individual posterior densities. Unlike other approaches, such as ICA or clustering, these images provide a truly probabilistic interpretation of a voxel's membership in a particular network.

References

1. Bellec, P., Rosa-Neto, P., Lyttelton, O., Benali, H., Evans, A.: Multi-level bootstrap analysis of stable clusters in resting-state fMRI. Neuroimage 51(3), 1126–1139 (2010)
2. Calhoun, V., Adali, T., Pearlson, G., Pekar, J.: Spatial and temporal independent component analysis of functional MRI data containing a pair of task-related waveforms. HBM 13(1), 43–53 (2001)
3. Descombes, X., Kruggel, F., Von Cramon, D.: Spatio-temporal fMRI analysis using Markov random fields. IEEE TMI 17(6), 1028–1039 (1998)
4. Liu, W., Zhu, P., Anderson, J.S., Yurgelun-Todd, D., Fletcher, P.T.: Spatial Regularization of Functional Connectivity Using High-Dimensional Markov Random Fields. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part II. LNCS, vol. 6362, pp. 363–370. Springer, Heidelberg (2010)
5. Ng, B., Abugharbieh, R., Hamarneh, G.: Group MRF for fMRI activation detection. In: CVPR 2010, pp. 2887–2894. IEEE (2010)
6. Ng, B., McKeown, M., Abugharbieh, R.: Group replicator dynamics: A novel groupwise evolutionary approach for sparse brain network detection. IEEE TMI 31(3), 576–585 (2012)
7. Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE PAMI 22(8), 888–905 (2000)
8. Van Den Heuvel, M., Mandl, R., Pol, H.: Normalized cut group clustering of resting-state FMRI data. PLoS One 3(4), e2001 (2008)
9. Varoquaux, G., Gramfort, A., Pedregosa, F., Michel, V., Thirion, B.: Multi-subject Dictionary Learning to Segment an Atlas of Brain Spontaneous Activity. In: Székely, G., Hahn, H.K. (eds.) IPMI 2011. LNCS, vol. 6801, pp. 562–573. Springer, Heidelberg (2011)

Metamorphic Geodesic Regression

Yi Hong[1], Sarang Joshi[3], Mar Sanchez[4],
Martin Styner[1], and Marc Niethammer[1,2]

[1] UNC-Chapel Hill
[2] Biomedical Research Imaging Center, UNC-Chapel Hill
[3] University of Utah
[4] Emory University

Abstract. We propose a *metamorphic* geodesic regression approach approximating spatial transformations for image time-series while simultaneously accounting for intensity changes. Such changes occur for example in magnetic resonance imaging (MRI) studies of the developing brain due to myelination. To simplify computations we propose an approximate metamorphic geodesic regression formulation that only requires pairwise computations of image metamorphoses. The approximated solution is an appropriately weighted average of initial momenta. To obtain initial momenta reliably, we develop a shooting method for image metamorphosis.

1 Introduction

To study aging, disease progression or brain development over time, longitudinal imaging studies are frequently used. Image registration is required if local structural changes are to be assessed. Registration methods that account for temporal dependencies in longitudinal imaging studies are recent, including generalizations of linear regression or splines for shapes [1,2] or images [3] and methods with general temporal smoothness penalties [4,5]. Changes in image intensities are generally not explicitly captured and instead accounted for by using image similarity measures which are insensitive to such changes. However, approaches accounting for intensity changes *after* registration exist [6].

We generalize linear regression to image time-series, capturing spatial and intensity changes *simultaneously*. This is achieved by a metamorphic regression formulation combining the dynamical systems formulation for geodesic regression for images [3] with image metamorphosis [7,8], similar to [9] for the large displacement diffeomorphic metric mapping (LDDMM) case. While several methods have been proposed to simultaneously capture image deformations and intensity changes for image registration [10,11,12] the metamorphosis approach [8] is most suitable here, because spatial deformations and intensity variations are described by a geodesic. This allows generalizing the concept of a regression line.

Sec. 2 reviews image metamorphosis and its relation to LDDMM. Sec. 3 derives optimality conditions to allow for a shooting solution to metamorphosis using an augmented Lagrangian approach [13]. Sec. 4 discusses first- and second-order adjoint solutions. Sec. 5 introduces metamorphic geodesic regression. Sec. 5

shows how an approximate solution can be obtained by appropriate averaging of the initial momenta of independent pair-wise metamorphosis solutions. We show results on synthetic and real longitudinal image sequences in Sec. 7. The paper concludes with a summary and an outlook on future work.

2   Metamorphosis

Starting from the dynamical systems formulation for LDDMM image registration

$$E(v) = \frac{1}{2} \int_0^1 \|v\|_L^2 \ dt + \frac{1}{\sigma^2} \|I(1) - I_1\|^2, \quad \text{s.t.} \ \ I_t + \nabla I^T v = 0, \ I(0) = I_0, \quad (1)$$

image metamorphosis allows exact matching of a target image I_1 by a warped and intensity-adjusted source image $I(1)$ by adding a control variable, q, which smoothly adjusts image intensities along streamlines. Here, $\sigma > 0$, v is a spatio-temporal velocity field and $\|v\|_L^2 = \langle Lv, Lv \rangle$, where L is a differential operator penalizing non-smooth velocities. The optimization problem changes to [8,7]

$$E(v, q) = \frac{1}{2} \int_0^1 \|v\|_L^2 + \rho \|q\|_Q^2 \ dt, \quad \text{s.t.} \ \ I_t + \nabla I^T v = q, \ I(0) = I_0, \ I(1) = I_1. \quad (2)$$

The inexact match of the final image is replaced by an exact matching, hence the energy value depends on the images to be matched only implicitly through the initial and final constraints; $\rho > 0$ controls the balance between intensity blending and spatial deformation. The solution to both minimization problems (1) and (2) is given by a geodesic, which is specified by its initial conditions. The initial conditions can be numerically computed through a shooting method.

3   Optimality Conditions for Shooting Metamorphosis

To derive the second order dynamical system required for a shooting method, we add the dynamical constraint through the momentum variable, p. Eq. 2 becomes

$$E(v, q, I, p) = \int_0^1 \frac{1}{2} \|v\|_L^2 + \frac{1}{2} \rho \|q\|_Q^2 + \langle p, I_t + \nabla I^T v - q \rangle \ dt, \quad \text{s.t.} \ \ I(0) = I_0, \ I(1) = I_1. \quad (3)$$

To simplify the numerical implementation we use an augmented Lagrangian approach [13] converting the optimization problem (3) to

$$E(v, q, I, p) = \int_0^1 \frac{1}{2} \|v\|_L^2 + \frac{1}{2} \rho \|q\|_Q^2 + \langle p, I_t + \nabla I^T v - q \rangle \ dt$$
$$- \langle r, I(1) - I_1 \rangle + \frac{\mu}{2} \|I(1) - I_1\|^2, \quad \text{s.t.} \ \ I(0) = I_0, \quad (4)$$

where $\mu > 0$ and r is the Lagrangian multiplier function for the final image-match constraint. The variation of Eq. 4 results in the optimality conditions

$$\begin{cases} I_t + \nabla I^T v & = \frac{1}{\rho}(Q^\dagger Q)^{-1}p, \ I(0) = I_0, \\ -p_t - div(pv) & = 0, \ p(1) = r - \mu(I(1) - I_1). \\ L^\dagger L v + p\nabla I & = 0, \end{cases} \tag{5}$$

The optimality conditions do not depend on q, since by optimality $q = \frac{1}{\rho}(Q^\dagger Q)^{-1}p$. Hence, the state for metamorphosis is identical to the state for LDDMM registration, (I, p), highlighting the tight coupling in metamorphosis between image deformation and intensity changes. The final state constraint $I(1) = I_1$ has been replaced by an augmented Lagrangian penalty function.

4   Shooting for Metamorphosis

The metamorphosis problem (2) has so far been addressed as a boundary value problem by relaxation approaches [14,8]. This approach hinders the formulation of the regression problem and assures geodesics at convergence only. We propose a shooting method instead. Since the final constraint has been successfully eliminated through the augmented Lagrangian approach, $\nabla_{p(0)}E$ can be computed using a first- or second-order adjoint method similarly as for LDDMM registration [15,16]. The numerical solution alternates between a descent step for $p(0)$ for fixed r, μ and (upon reasonable convergence) an update step

$$r^{(k+1)} = r^{(k)} - \mu^{(k)}(I(1) - I_1).$$

The penalty parameter μ is increased as desired such that $\mu^{(k+1)} > \mu^{(k)}$. Numerically, we solve all equations by discretizing time, assuming v and p to be piece-wise constant in a time-interval. We solve transport equations and scalar conservation laws by propagating maps [17] to limit numerical dissipation.

4.1   First-Order Adjoint Method

Following [16], we can compute $\nabla_{v(0)}E$ by realizing that the Hilbert gradient is

$$\nabla_{v(0)}E = v(0) + K * (p(0)\nabla I(0)),$$

where $K = (L^\dagger L)^{-1}$. Therefore based on the adjoint solution method [17,18]

$$\nabla_{v(0)}E = v(0) + K * (\hat{p}(0)\nabla I(0)) = v(0) + K * (|D\Phi|\hat{p}(1) \circ \Phi\nabla I(0)), \tag{6}$$

where Φ is the map from $t = 1$ to $t = 0$ given the current estimate of the velocity field $v(x, t)$ and $\hat{p}(1) = r - \mu(I(1) - I_1)$ with $I(1) = I_0 \circ \Phi^{-1}$. Storage of the time-dependent velocity fields is not required as both Φ and Φ^{-1} can be computed and stored during a forward (shooting) sweep. Instead of performing the

gradient descent on $v(0)$ it is beneficial to compute it directly with respect to $p(0)$ since this avoids unnecessary matrix computation. Since at $t = 0$: $-(L^\dagger L)\delta v(0) = \delta p(0)\nabla I(0)$, it follows from Eq. 6 that

$$\nabla_{p(0)} E = p(0) - \hat{p}(0) = p(0) - |D\Phi|(r - \mu(I(1) - I_1)) \circ \Phi.$$

4.2   Second-Order Adjoint Method

The energy can be rewritten in initial value form (wrt. $(I(0), p(0))$) as

$$E = \frac{1}{2}\langle p(0)\nabla I(0), K * (p(0)\nabla I(0))\rangle + \frac{1}{2\rho}\langle (Q^\dagger Q)^{-1}p(0), p(0)\rangle$$
$$- \langle r, I(1) - I_1\rangle + \frac{\mu}{2}\|I(1) - I_1\|^2, \quad \text{s.t. Eq. (5) holds.}$$

At optimality, the state equations (5) and

$$\begin{cases} -\lambda^I_t - div(v\lambda^I) & = div(pK * \lambda^v), \\ -\lambda^p_t - v^T\nabla\lambda^p & = -\nabla I^T K * \lambda^v + \frac{1}{\rho}(Q^\dagger Q)^{-1}\lambda^I, \\ \lambda^I\nabla I - p\nabla\lambda^p + \lambda^v & = 0, \end{cases}$$

hold, with final conditions: $\lambda^p(1) = 0$; $\lambda^I(1) = r - \mu(I(1) - I_1)$. The gradient is

$$\nabla_{p(0)} E = -\lambda^p(0) + \nabla I(0)^T K * (p(0)\nabla I(0)) + \frac{1}{\rho}(Q^\dagger Q)^{-1}p(0).$$

The dynamic equations and the gradient are only slightly changed from the LDDMM registration [15] when following the augmented Lagrangian approach.

5   Metamorphic Geodesic Regression

Our goal is the estimation of a regression geodesic (under the geodesic equations for metamorphosis) wrt. a set of measurement images $\{I_i\}$ by minimizing

$$E = \frac{1}{2}\langle m(t_0), K * m(t_0)\rangle + \frac{1}{2\rho}\langle (Q^\dagger Q)^{-1}p(t_0), p(t_0)\rangle + \frac{1}{\sigma^2}\sum_{i=1}^{N} \text{Sim}(I(t_i), I_i) \quad (7)$$

such that Eq. (5) holds. Here, $\sigma > 0$ balances the influence of the change of the regression geodesic with respect to the measurements, $m(t_0) = p(t_0)\nabla I(t_0)$ and Sim denotes an image similarity measure. A solution scheme with respect to $(I(t_0), p(t_0))$ can be obtained following the derivations for geodesic regression [3]. Such a solution requires the integration of the state equation as well as the second-order adjoint. Further, for metamorphosis it is sensible to also define $\text{Sim}(I(t_i), I_i)$ based on the squared distance induced by the solution of the metamorphosis problem between $I(t_i)$ and I_i. Since no closed-form solutions for these distances are computable in the image-valued case an iterative

solution method is required which would in turn require the underlying solution of metamorphosis problems for each measurements at each iteration. This is costly.

6   Approximated Metamorphic Geodesic Regression

To simplify the solution of metamorphic geodesic regression (7), we approximate the distance between two images I_1, I_2 wrt. a base image I_b at time t as

$$
Sim(I_1, I_2) = d^2(I_1, I_2) \approx t^2 \frac{1}{2} \langle m_1(0) - m_2(0), K * (m_1(0) - m_2(0)) \rangle
$$
$$
+ t^2 \frac{1}{2\rho} \langle (Q^\dagger Q)^{-1}(p_1(0) - p_2(0)), p_1(0) - p_2(0) \rangle, \quad (8)
$$

where $p_1(0)$ and $p_2(0)$ are the initial momenta for I_1 and I_2 wrt. the base image I_b (i.e., the initial momenta obtained by solving the metamorphosis problem between I_b and I_1 as well as for I_b and I_2 respectively) and $m_1(0) = p_1(0)\nabla I_b$, $m_2(0) = p_2(0)\nabla I_b$. This can be seen as a tangent space approximation for metamorphosis. The squared time-dependence emerges because the individual difference terms are linear in time.

We assume that the initial image $I(t_0)$ on the regression geodesic is known. This is a simplifying assumption, which is meaningful for example for growth modeling wrt. a given base image[1]. Substituting into Eq. (7) yields

$$
E(p(t_0)) = \frac{1}{2}\langle m(t_0), K * m(t_0) \rangle + \frac{1}{2\rho}\langle (Q^\dagger Q)^{-1}p(t_0), p(t_0) \rangle
$$
$$
+ \frac{1}{\sigma^2}\sum_{i=1}^{N}\frac{1}{2}(\Delta t_i)^2 \langle m(t_0) - m_i, K*(m(t_0) - m_i) \rangle + \frac{1}{2\rho}(\Delta t_i)^2 \langle (Q^\dagger Q)^{-1}(p(t_0) - p_i), p(t_0) - p_i \rangle.
$$

Here, $m(t_0) = p(t_0)\nabla I(t_0)$, $\Delta t_i = t_i - t_0$, $m_i = p_i \nabla I(t_0)$ and p_i is the initial momentum for the metamorphosis solution between $I(t_0)$ and I_i. For a given $I(t_0)$, the p_i can be independently computed. The approximated energy only depends on the initial momentum $p(t_0)$. The energy variation yields the condition

$$
R[(1 + \frac{1}{\sigma^2}\sum_{i=1}^{N}(\Delta t_i)^2)p(t_0)] = R[\frac{1}{\sigma^2}\sum_{i=1}^{N}(\Delta t_i)^2 p_i],
$$

where the operator R is $R[p] := \nabla I(t_0)^T K * (\nabla I(t_0)p) + \frac{1}{\rho}(Q^\dagger Q)^{-1}p$. Since $K = (L^\dagger L)^{-1}$ and $\rho > 0$ this operator is invertible and therefore

---

[1] Ideally one would like to construct an image on the geodesic given all the measurement images and then perform all computations with respect to it. For the linear regression model the point defined by the mean in time and the measurements, $(\bar{t}, \bar{y})$, is on the regression line. If such a relation exists for metamorphic geodesic regression, e.g., some form of unbiased mean with similar properties, remains to be determined.

$$p(t_0) = \frac{\frac{1}{\sigma^2} \sum_{i=1}^{N} (\Delta t_i)^2 p_i}{1 + \frac{1}{\sigma^2} \sum_{i=1}^{N} (\Delta t_i)^2} \overset{\sigma \to 0}{\approx} \frac{\sum_{i=1}^{N} (\Delta t_i)^2 p_i}{\sum_{i=1}^{N} (\Delta t_i)^2}.$$

The last approximation is sensible since typically $\sigma \ll 1$. It recovers the metamorphosis solution if there is only one measurement image and the base image.

7   Experimental Results

7.1   Simulated Examples

In Fig. 1, four images (32×32, spacing 0.04) are synthesized to simulate the movement of a bull's eye. The outside white loop of the eye shrinks with no intensity changes, while the inside circle grows at a constant speed and its intensity changes from white to gray. The images are at time instants 0, 10, 20, 30 and we chose the first one as the base image. Eight Gaussian kernels [19] are used for K: $\{K_{0.5}, K_{0.4}, K_{0.3}, K_{0.25}, K_{0.2}, K_{0.15}, K_{0.1}, K_{0.05}\}$; $\rho = 0.75$. The result confirms that the spatial transformation and intensity changes are captured simultaneously. The dark solid circle at the center of the average momentum of Fig. 1 indicates that the intensity of the inside circle will decrease gradually. The white loop outside of the dark area captures the growth of the inside circle.

Fig. 2 shows a square (64×64; spacing 0.02) moving from left to right at a uniform speed with gradually decreasing intensity. Measurements are at 0, 10, 20, 30, 40. We used a multi-Gaussian kernel K with $\{K_{1.0}, K_{0.75}, K_{0.5}, K_{0.4}, K_{0.3}, K_{0.2}, K_{0.1}\}$ and set $\rho = 5.0$. Metamorphic regression successfully captures the spatial transformation and the intensity changes of the square even when adding vertical oscillations. As expected, metamorphic regression eliminates the oscillations while capturing the intensity change and the movement to the right. We see from the time-weighted average of the initial momenta that intensity changes are controlled by the values inside the square region (dark: decreasing intensity; bright: increasing intensity). The spatial transformations are mainly controlled by the momenta on the edges of the square.
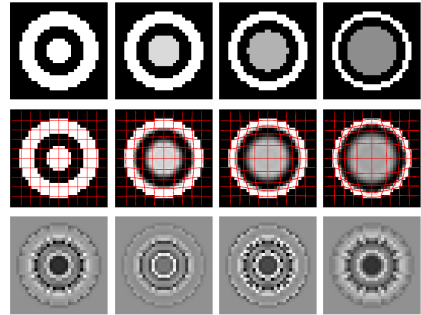

Fig. 1. Bull's eye metamorphic regression experiment. Measurement images (top row). Metamorphic regression result (middle row) and momenta (bottom row). The first image is chosen as base image. Momenta images: left: time-weighted average of the initial momenta; right: momenta of the measurement images with respect to the base image.
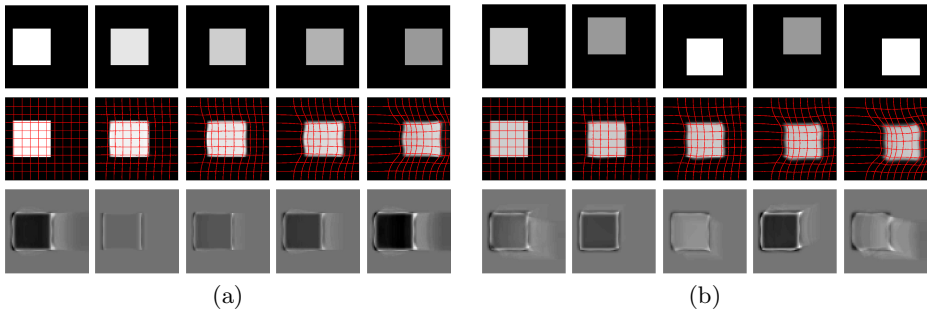
Fig. 2. Square metamorphic regression experiment. (a) moving square with decreasing intensities and no oscillations during movement; (b) moving and oscillating square with alternating intensities. For both cases, the base image is the first one. Top row: measurement images, middle row: metamorphic regression results, bottom row: momenta images (left: time-weighted average of the initial momenta, to the right: momenta of the measurement images with respect to the base image).

7.2   Real Images

Fig. 3 shows two representative longitudinal MRI time-series (300×250 with spacing 0.2734) of nine macaque monkeys at age 3, 6, and 12 months. Some subjects have no visible myelination in the anterior parts of the brain at 3 months (top left), while others show substantial myelination (bottom left). Here, we use metamorphic geodesic regression not for an individual longitudinal image set, but for all nine monkeys and all time-points simultaneously. We use an unbiased atlas for images at 12 months as the base image. Metamorphic geodesic regression is applied over the remaining 18 images at 3 and 6 months. We use a multi-Gaussian kernel, K, with $\{K_{40}, K_{20}, K_{15}, K_{10}, K_5, K_{2.5}\}$; $\rho = 500$.
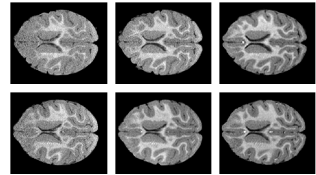

Fig. 3. Representative datasets at 3, 6 and 12 months (left to right)
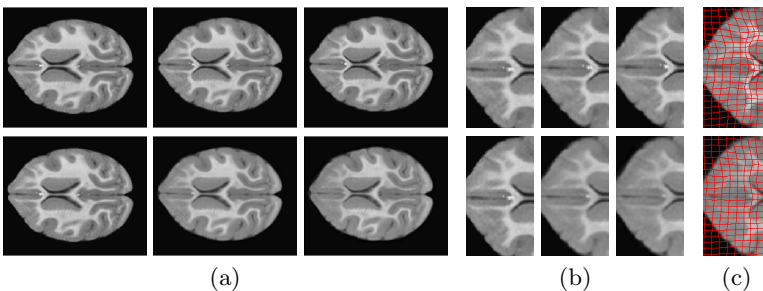

Fig. 4. Regression results for monkey data: LDDMM (top) metamorphosis (bottom). (a) Images on geodesic at 12, 6, 3 months; (b) Zoom in for images on geodesic at 12, 6, 3 months; (c) Zoom in for images at 3 months to illustrate spatial deformation.

Fig. 4 shows regression results for the simple metamorphic model and for its LDDMM version [9] which cannot capture intensity changes. The metamorphic regression geodesic captures intensity changes of the brain well (increase in white matter intensity with age caused by myelination) while capturing spatial deformations, most notably a subtle expansion of the ventricles.

8   Discussion and Conclusions

We proposed metamorphic geodesic regression for image time-series which simultaneously captures spatial deformations and intensity changes. For efficient computations we use a tangent space approximation with respect to a chosen base-image. Solutions can be computed by solving pairwise metamorphosis problems through a shooting approach. Future work will address the properties of the approximation, alternative models of intensity change and the trade-off between spatial deformation and change in image intensities.

References

1. Fletcher, T.: Geodesic regression on Riemannian manifolds. In: MICCAI Workshop on Mathematical Foundations of Computational Anatomy, pp. 75–86 (2011)
2. Trouvé, A., Vialard, F.: A second-order model for time-dependent data interpolation: Splines on shape spaces. In: Workshop STIA-MICCAI (2010)
3. Niethammer, M., Huang, Y., Vialard, F.-X.: Geodesic Regression for Image Time-Series. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part II. LNCS, vol. 6892, pp. 655–662. Springer, Heidelberg (2011)
4. Durrleman, S., Pennec, X., Trouvé, A., Gerig, G., Ayache, N.: Spatiotemporal Atlas Estimation for Developmental Delay Detection in Longitudinal Datasets. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009, Part I. LNCS, vol. 5761, pp. 297–304. Springer, Heidelberg (2009)
5. Fishbaugh, J., Durrleman, S., Gerig, G.: Estimation of Smooth Growth Trajectories with Controlled Acceleration from Time Series Shape Data. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part II. LNCS, vol. 6892, pp. 401–408. Springer, Heidelberg (2011)
6. Rohlfing, T., Sullivan, E.V., Pfefferbaum, A.: Regression Models of Atlas Appearance. In: Prince, J.L., Pham, D.L., Myers, K.J. (eds.) IPMI 2009. LNCS, vol. 5636, pp. 151–162. Springer, Heidelberg (2009)
7. Holm, D.D., Trouvé, A., Younes, L.: The Euler-Poincaré theory of metamorphosis. Quarterly of Applied Mathematics 67, 661–685 (2009)
8. Miller, M.I., Younes, L.: Group actions, homeomorphisms, and matching: A general framework. International Journal of Computer Vision 41, 61–84 (2001)

9. Hong, Y., Shi, Y., Styner, M., Sanchez, M., Niethammer, M.:Simple Geodesic Regression for Image Time-Series. In: Dawant, B.M., Christensen, G.E., Fitzpatrick, J.M., Rueckert, D. (eds.) WBIR 2012. LNCS, vol. 7359, pp. 11–20. Springer, Heidelberg (2012)

10. Gupta, S.N., Prince, J.L.: On variable brightness optical flow for tagged MRI. In: IPMI, pp. 323–334 (1995)

11. Friston, K., Ashburner, J., Frith, C., Poline, J., Heather, J., Frackowiak, R.: Spatial registration and normalization of images. Hum. Brain Mapp. 3(3), 165–189 (1995)

12. Periaswamy, S., Farid, H.: Elastic registration in the presence of intensity variations. IEEE Trans. Med. Imaging 22(7), 865–874 (2003)

13. Nocedal, J., Wright, S.: Numerical optimization. Springer (1999)

14. Garcin, L., Younes, L.: Geodesic Image Matching: A Wavelet Based Energy Minimization Scheme. In: Rangarajan, A., Vemuri, B.C., Yuille, A.L. (eds.) EMMCVPR 2005. LNCS, vol. 3757, pp. 349–364. Springer, Heidelberg (2005)

15. Vialard, F., Risser, L., Rueckert, D., Cotter, C.: Diffeomorphic 3D image registration via geodesic shooting using an efficient adjoint calculation. International Journal of Computer Vision 97(2), 229–241 (2012)

16. Ashburner, J., Friston, K.: Diffeomorphic registration using geodesic shooting and Gauss-Newton optimization. Neuroimage 55(3), 954–967 (2011)

17. Beg, M., Miller, M., Trouvé, A., Younes, L.: Computing large deformation metric mappings via geodesic flows of diffeomorphisms. IJCV 61(2), 139–157 (2005)

18. Hart, G.L., Zach, C., Niethammer, M.: An optimal control approach for deformable registration. In: MMBIA, pp. 9–16 (2009)

19. Risser, L., Vialard, F., Wolz, R., Murgasova, M., Holm, D., Rueckert, D.: Simultaneous multiscale registration using large deformation diffeomorphic metric mapping. IEEE Transactions on Medical Imaging 30(10), 1746–1759 (2011)

Eigenanatomy Improves Detection Power
for Longitudinal Cortical Change

Brian Avants, Paramveer Dhillon, Benjamin M. Kandel, Philip A. Cook,
Corey T. McMillan[1], Murray Grossman[1], and James C. Gee

[1] Departments of Radiology and Neurology
University of Pennsylvania, Philadelphia, PA 19104

Abstract. We contribute a novel and interpretable dimensionality re-
duction strategy, *eigenanatomy*, that is tuned for neuroimaging data.
The method approximates the eigendecomposition of an image set with
basis functions (the eigenanatomy vectors) that are *sparse*, *unsigned*
and are *anatomically clustered*. We employ the eigenanatomy vectors
as anatomical predictors to improve detection power in morphometry.
Standard voxel-based morphometry (VBM) analyzes imaging data
voxel-by-voxel—and follows this with cluster-based or voxel-wise mul-
tiple comparisons correction methods to determine significance. Eige-
nanatomy reverses the standard order of operations by first clustering
the voxel data and then using standard linear regression in this reduced
dimensionality space. As with traditional region-of-interest (ROI) analy-
sis, this strategy can greatly improve detection power. Our results show
that eigenanatomy provides a principled objective function that leads
to localized, data-driven regions of interest. These regions improve our
ability to quantify biologically plausible rates of cortical change in two
distinct forms of neurodegeneration. We detail the algorithm and show
experimental evidence of its efficacy.

1   Introduction

In machine learning, interpretable data decompositions are termed "parts-based
representations" because they transform unstructured data into interpretable
pieces [1–3]. Recent work in machine learning points to the fact that exploiting
problem-specific information can improve parts-based representations [4–6]. Un-
informed, generic matrix decomposition methods, e.g. standard principal com-
ponent analysis (PCA), may be difficult to interpret because the solutions will
produce vectors that are everywhere non-zero, i.e. involve the whole brain rather
than its parts. Sparse methods have sought to resolve this issue [2, 3, 7–10]. How-
ever, these recent sparse multivariate methods are anatomically uninformed.

In this work, we employ a novel data-driven framework, related to the methods
above, to delineate cortical networks wherein longitudinal atrophy patterns in
Alzheimer's disease (AD) and frontotemporal lobar degeneration (FTLD) differ
from controls. Our novel image processing framework is open-source, unbiased
with respect to registration and segmentation [11, 12] and is formulated spa-
tiotemporally, as described in [13]. At the statistical level, our method is unbiased

in that it uses the intrinsic covariation of the dataset to parcellate the cortex into coherent regions and, in this reduced space, we gain sensitivity over full voxel-wise testing with voxel-based morphometry (VBM) [14]. Dimensionality reduction is critical for datasets that are relatively small and yet which quantify valuable and difficult to collect measurements in uncommon populations of subjects, as in FTLD. Critically, our novel dimensionality reduction method provides an objective function that optimally maps the classical singular value decomposition (SVD) eigenvectors that are both *signed* and *global* into *spatially localized, sparse, unsigned* pseudo-eigenvectors (or eigenanatomy). To our knowledge, the specific objective function used in eigenanatomy is the first to formulate an unsigned sparse decomposition with explicit guidance by the SVD solution and with anatomically informed regularization. We apply this framework in a cohort that is diagnosed by CSF biofluid biomarkers with high specificity and sensitivity [15]. This well-defined cohort provides an excellent test-bed for this new algorithm as we expect to identify specific patterns of cortical atrophy within regions known to be affected in FTLD and AD. At the same time, the relatively small cohort and challenges of longitudinal mapping may make this quantification difficult to achieve with VBM. We also show that eigenanatomy produces more powerful predictors when compared to related methods: VBM, classic SVD and penalized matrix decomposition (PMD) [16]. PMD is freely available and provides a sparse approach to PCA.

2 Methods

The class of methods encompassing non-negative matrix factorization (NMF) [1, 3, 17, 18], sparse principal components analysis (SPCA) [2, 19, 16, 20] and singular value decomposition [21] form the basis for the approach proposed here. More formally, define a $n \times p$ (rows by columns) matrix $\mathbf{X}$ where each row derives from an observed subject image such that the collection of images is given as vectors $\{x_1, ..., x_n\}$ with each vector x_i containing p entries. First, we denote each eigenanatomy component (a pseudo-eigenvector) as $\mathbf{v}_i$ where i is ordered such that each eigenanatomy from $\mathbf{v}_1$ to $\mathbf{v}_m$ provides a decreasing contribution to the variance of matrix $\mathbf{X}$. We define eigenanatomy pseudo-eigenvectors as *sparse*. Sparseness means that some entries in $\mathbf{v}_j$ will be zero.

The classic singular value decomposition (SVD) may be used to reduce the dimensionality of this data by decomposing the dataset into the eigenvectors of $C_p = \mathbf{X}^T \mathbf{X}$ and $C_n = \mathbf{X}\mathbf{X}^T$ (the right and left singular vectors, respectively). The relationship between a C_p and a C_n eigenvector is given by $\mathbf{X}\mathbf{v}^p = \mathbf{v}^n$ and $\mathbf{X}^T\mathbf{v}^n = \mathbf{v}^p$. The n eigenvectors from C_p and C_n may be used to reconstruct the matrix $\mathbf{X}$ by $\sum_i \mathbf{v}_i^n \otimes \mathbf{v}_i^p \lambda_i$ where the λ_i denotes the ith eigenvalue and $\otimes$ the outer product. We now detail the algorithm that comprises the main contribution of this work and which is available as part of the *sccan* program within Advanced Normalization Tools (ANTs) [13].

Our goal is to approximate the matrix $\mathbf{X}$ with its right and left singular vectors but where the right singular vector is *sparse*. We might, then, minimize:

$$\|\mathbf{X} - \sum_i \mathbf{v}_i^n \otimes \mathbf{v}_i^{sp} \lambda_i\|^2 \tag{1}$$

where the $\mathbf{v}_i^{sp}$ denotes the ith *sparse* right singular vector. It is known that the optimal value for $\mathbf{v}_i^{sp}$ is exactly $\mathbf{X}^T \mathbf{v}^n$ (if we relax sparseness constraints). Using the fact that $\mathbf{X}\mathbf{v}^p = \mathbf{v}^n$ and $\mathbf{X}^T \mathbf{v}^n = \mathbf{v}^p$, we therefore reformulate the objective in a slightly simpler form and seek to directly find $\mathbf{v}_i^{sp}$ by optimizing:

$$\underset{\mathbf{v}_i^{sp}}{\arg\min} \ \|\mathbf{X}\mathbf{v}_i^{sp} - \mathbf{v}_i^n\|^2. \tag{2}$$

This optimization problem is quadratic without sparseness constraints and easily solved by conjugate gradient through the normal equations $\|\mathbf{X}^T\mathbf{X}\mathbf{v}_i^{sp} - \mathbf{X}^T\mathbf{v}_i^n\|^2$.

Now, note that the vector $\mathbf{X}^T\mathbf{v}_i^n = \mathbf{v}_i^p$ might have both positive and negative values. As with non-negative matrix factorization, we seek a decomposition that is unsigned. Thus, an optimal solution that minimizes $\|\mathbf{X}\mathbf{v}_i^{sp} - \mathbf{v}_i^n\|^2$ will need to model both signs. We therefore make a second adjustment by modeling the positive and negative components of $\mathbf{v}_i^p$ separately.

Each eigenvector may be written in an expanded form via the use of indicator functions which are diagonal matrices with binary entries. For instance, if $\mathbf{v}$ contains entries $[\ -2,\ -1,\ 0,\ 1,\ 2\ ]$, then the positive indicator function is $I^+ = [\ 0,\ 0,\ 0,\ 1,\ 1\ ]$ and the negative indicator function is $I^- = [\ 1,\ 1,\ 0,\ 0,\ 0\ ]$. $\mathbf{v}$ may then be expressed as $\mathbf{v} = I^+\mathbf{v} + I^-\mathbf{v} = \mathbf{v}^+ + \mathbf{v}^-$. We use these indicator functions to separate the positive and negative components of our objective such that the optimization in equation 2 becomes,

$$\underset{\mathbf{v}_i^{sp+},\ \mathbf{v}_i^{sp-}}{\arg\min} \ \|\mathbf{X}^T\mathbf{X}\mathbf{v}_i^{sp+} - \mathbf{v}_i^{p+}\|^2 + \|\mathbf{X}^T\mathbf{X}\mathbf{v}_i^{sp-} - \mathbf{v}_i^{p-}\|^2 \tag{3}$$

This minimization problem forms the basis for our novel approach to computing eigenanatomy, i.e. anatomically localized approximations to the eigenvectors of an anatomical imaging dataset. Derivation of sparse eigenanatomy is shown in Figure 1.

As noted above, we seek sparse and interpretable solutions. We define a sparse vector as one which minimizes a l_0 or l_1 penalty term i.e. has either a user-specified number of non-zero entries (l_0) or absolute sum (l_1). Although the l_1 penalty has advantages [2], we use the l_0 penalty because it specifies the fraction of the vector that is allowed to be non-zero. The sparseness restriction is therefore easily interpreted by users of the eigenanatomy method. Eigenanatomy seeks to identify sparse functions $\mathbf{v}_i^{sp+}$ and $\mathbf{v}_i^{sp-}$ that closely approximate the eigenvectors in n-space, i.e. $\mathbf{v}_n^i = \mathbf{X}\mathbf{v}_i^p$. The objective function, again employing the normal equations, is then:

$$\underset{\mathbf{v}_i}{\arg\min} \ \sum_{i=1}^{n} \| \ C_p\mathbf{v}_i^{sp+} - \mathbf{v}_i^{p+} \ \|^2 + \| \ C_p\mathbf{v}_i^{sp-} - \mathbf{v}_i^{p-} \ \|^2 \tag{4}$$

$$\text{subject to:} \quad \|\mathbf{v}_i^{sp+}\|_0 = \|\mathbf{v}_i^{sp-}\|_0 = \gamma,$$

Algorithm 1. SNLCG optimization for eigenanatomy.

Input $\mathbf{X}$, the eigenvectors of $\mathbf{X}\mathbf{X}^T$ and γ, the sparseness parameter.
for all $\mathbf{v}_i \in \{\mathbf{v}_1, \ldots, \mathbf{v}_{n-1}\}$ **do**
    $\mathbf{v}_i \leftarrow \mathbf{X}^T \mathbf{v}_i^n$ $\triangleright$ Get the p-space eigenvector from the $\mathbf{v}_i^n$.
    Compute $\mathbf{v}_i^+$ and $\mathbf{v}_i^-$. $\triangleright$ Find the $+$ and $-$ representation of $\mathbf{v}_i$.
    $\mathbf{v}_i^{s+} \leftarrow \mathbf{v}_i^{s-} \leftarrow \frac{1}{p}$ $\triangleright$ Initialize the sparse $+$ and $-$ vectors.
    $\mathbf{v}_i^{s+} \leftarrow \text{SNLCG}(\mathbf{X}, \mathbf{v}_i^{s+}, \mathbf{v}_i^+, \gamma)$ $\triangleright$ SNLCG $+$ minimization.
    $\mathbf{v}_i^{s-} \leftarrow \text{SNLCG}(\mathbf{X}, \mathbf{v}_i^{s-}, \mathbf{v}_i^-, \gamma)$ $\triangleright$ SNLCG $-$ minimization.
    $\mathbf{v}_i^{s-} \leftarrow \mathbf{v}_i^{s-} * (-1)$ $\triangleright$ Reset $\mathbf{v}_i^{s-}$ to be positive.
end for

---

where γ defines the desired level of sparseness for each eigenanatomy vector. Eigenanatomy therefore produces $2*n$ sparse pseudo-eigenvectors whose product with $\mathbf{X}$ may be used a predictors in standard linear regression. Importantly, because these vectors are unsigned, they may be interpreted as *weighted averages of the input data.*

Sparseness can be enforced by a soft-thresholding algorithm as in [2, 16]. We denote this function as $S(\mathbf{v}, \gamma)$ and (in an *adhoc* manner) allow it to also reject isolated voxels of the eigenanatomy vector that are non-contiguous (i.e. we provide a cluster threshold as in VBM). Minimization problems involving the l_0 penalty are *np*-hard. The relaxed form of this objective function (i.e. without the sparseness constraint) is purely quadratic and can easily be solved by a conjugate gradient method. Thus, we propose a new sparse, nonlinear conjugate gradient (SNLCG) method as a minimization procedure for the eigenanatomy objective function to deal with the nonlinearities induced by the S function and l_0 constraint. The additional advantage of SNLCG is that its solutions approach the quadratic minimum as sparseness constraints are relaxed. We detail the minimization algorithms for the eigenanatomy objective function (equation 4) in algorithms 1 and 2. The algorithms are also available in an open-source **R** package for free use and within the ANTs toolkit.

3 Results

For all uses of eigenanatomy below, we set the sparseness parameter, γ, to select 5% of the voxels in the cortex. We choose 5% because this provides interpretable clusters of regions in the cortex and yet still allows a reasonable reconstruction of the original data matrix, as shown in Figure 1.

Reconstruction Error: We quantify the ability of the eigenanatomy algorithm to reconstruct the original dataset from sparse eigenvectors using equation 1. As a baseline, we compare the full eigenanatomy solution to the reconstruction given by applying the soft-threshold function S directly to the SVD-derived
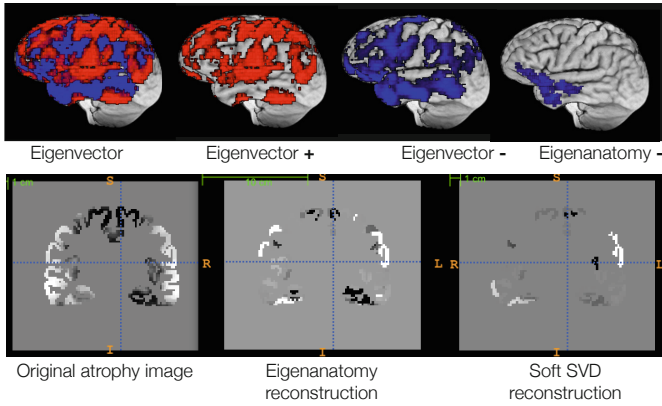
Fig. 1. The eigenanatomy basis functions: The original eigenvector (far left) has both positive and negative components. These are separated into the positive and negative vector components (middle figures). The sparse eigenanatomy approximation to $\mathbf{v}^-$ is shown at far right. Because the entries of $\mathbf{v}^{s-}$ are either zero or negative, the sign of $\mathbf{v}^{s-}$ can be changed to positive. Thus, $\mathbf{v}^{s-}$ is an interpretable measurement of the data and provides a weighted average of the original signal. Ultimately, the weighted average of the imaging data provided by $\mathbf{v}^{s-}$ is used as a predictor in regression. The same is done with $\mathbf{v}^{s+}$. In the lower portion of the figure, we see reconstruction results from the eigenanatomy method—see the results section for more explanation.

vectors $\mathbf{v}_i^p$ without further optimization. We call this "soft-SVD". We also compare reconstruction error to the eigenanatomy algorithm run with a restriction on the number of iterations in the SNLCG sub-algorithm. These experiments show that the full eigenanatomy algorithm run until convergence (error = 1.251) improves upon both the soft-SVD solution (error = 1.292) and the limited iteration eigenanatomy solution (error = 1.279). Error is measured by the frobenius norm taken between the original matrix and the reconstructed matrix.

Neuroimaging Data: Our cohort consists of 61 participants, including 15 patients with AD (7 females), 23 patients with FTLD (14 females), and 23 controls (13 females). All patients were clinically diagnosed by a board-certified neurologist and cerebrospinal fluid confirmation of the underlying pathology was obtained [15]. No significant difference exist between disease duration, age or education in the patient or control groups. Two 3.0T MPRAGE T1-weighted magnetic resonance images were obtained for each subject. The FTLD group (time interval 1.12 years +/- 0.28) trended ($p<0.081$) towards having a reduced interval when compared to elderly controls (time interval 1.29+/-0.36 years), as did the AD group (time interval 1.13+/-0.25 years, $p<0.12$). The interval between scans was therefore factored out as a nuisance variable.

Detection Power in Comparison to VBM, SVD and PMD: In this section, we employ eigenanatomy to compare the ability to detect group differences in cortical atrophy rate between FTLD subjects and controls as well as AD

Algorithm 2. SNLCG sub-algorithm for eigenanatomy.

Input $\mathbf{X}, \mathbf{v}_i^-, \mathbf{v}_i^{s-}, \gamma$.

$\mathbf{b} \leftarrow \mathbf{v}_i^-$

$\mathbf{x}_k \leftarrow \mathbf{v}_i^{s-}$

$\mathbf{r}_k \leftarrow (\; \mathbf{b} \; - \; \mathbf{X}^T(\; \mathbf{X}\,\mathbf{x}_k\;))$ ▷ Use the gradient of the quadratic term.

$\mathbf{p}_k \leftarrow \mathbf{r}_k$

$\Delta E \leftarrow \infty$

while $\Delta E > 0$ **do**

    $\alpha_k \leftarrow \langle\; \mathbf{p}_k,\; \mathbf{X}^T\,\mathbf{X}\,\mathbf{x}_k\; \rangle$ ▷ $\langle \cdot, \cdot \rangle$ denotes inner product.

    $\mathbf{x}_{k+1} \leftarrow \mathbf{x}_k + \alpha_k \mathbf{p}_k$

    $\mathbf{x}_{k+1} \leftarrow S(\mathbf{x}_{k+1}, \gamma)$

      ▷ Project $\mathbf{x}_{k+1}$ to the sparse solution space given by the objective function.

    $\mathbf{r}_{k+1} \leftarrow S(\; \mathbf{b} \; - \; \mathbf{X}^T(\; \mathbf{X}\,\mathbf{x}_{k+1}\;) , \gamma)$ ▷ Use the gradient of the quadratic term.

      ▷ Project $\mathbf{r}_{k+1}$ to the sparse solution space given by the objective function.

    $\beta_k \leftarrow \|\mathbf{r}_{k+1}\|^2 / \|\mathbf{r}_k\|^2$ ▷ Standard conjugate gradient definitions below.

    $p_{k+1} = r_{k+1} + \beta_k * p_k$

    $\Delta E \leftarrow \|\mathbf{r}_k\| - \|\mathbf{r}_{k+1}\|$

    $\mathbf{r}_k \leftarrow \mathbf{r}_{k+1}; \mathbf{x}_k \leftarrow \mathbf{x}_{k+1}; \mathbf{p}_k \leftarrow \mathbf{p}_{k+1}$.

end while

---

subjects and controls. This analysis shows specificity of the approach and biological plausibility in two different neurodegenerative disorders.

We passed the same input dataset to all methods. The data consisted of unbiased voxel-wise measures of annualized atrophy rate in the cortex of all patients normalized to a group template, as described elsewhere [13]. The regression model employed for all methods is summarized (in **R** syntax) as: *atrophy-rate* $\approx 1+$ diagnosis + education + interval-between-images + disease-duration + gender. "Diagnosis" is the predictor of interest i.e. we test whether the presence of disease predicts atrophy rate given the presence of the covariates. The *atrophy-rate* is either a vector of voxel-wise measures or a basis function projection against the original atrophy rate image matrix $\mathbf{X}$ i.e. $\mathbf{X}\mathbf{v}_i^p$. The latter case is used for classic SVD, PMD, soft-SVD in addition to eigenanatomy. We define significance as a q-value < 0.05 where a q-value is a false discovery rate corrected p-value. For PMD and SVD, we tested n different atrophy rates (one for each eigenvector) where n is number of subjects. For eigenanatomy and soft-SVD, we used $2n$ predictors as suggested by design in the algorithm. For VBM, we tested all 50,194 voxels (the number of columns in $\mathbf{X}$).

For the FTLD subjects and the AD subjects, when classic SVD projections were used as measures of atrophy rate, *no* significant predictors emerged. The same is true for the PMD method—although we note the caveat that, potentially, a more exhaustive parameter search may have resulted in better PMD results. Both univariate VBM and soft-SVD identify significant effects although the minimum q-value and extensiveness of both soft-SVD and VBM are far less than eigenanatomy. This is particularly true for the FTLD subjects for which VBM only produces a small 20 voxel cluster that survives correction. The results
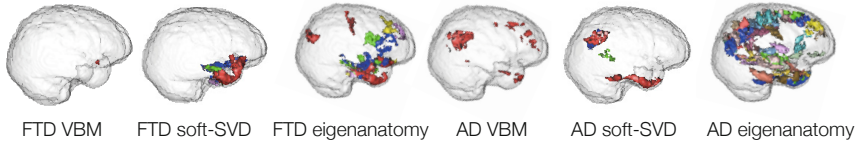
FTD VBM   FTD soft-SVD   FTD eigenanatomy   AD VBM   AD soft-SVD   AD eigenanatomy

Fig. 2. Statistical comparison: Eigenanatomy detects the most effects and also with smallest p-values (FTLD: q < 0.0015, AD: q < 0.002) versus the next best soft-SVD (FTLD: q < 0.0035, AD: q < 0.009). Different colors represent different eigenvectors / predictors. Note that multiple disjoint, but related, voxel clusters (i.e. a network) may be involved in an eigenanatomy vector. VBM detects 20 voxels in FTLD (q < 0.015). In AD, univariate results are more robust, likely due to the widespread nature of AD atrophy, and 1382 significant voxels (q < 0.0270) were detected.

are summarized further in Figure 2. Detailed clinical interpretation is beyond the scope of the paper but the results, overall, are coincident with what is known about these disorders. In particular, the largest atrophy rate in FTLD was in right orbitofrontal cortex. For AD, this region was in the precuneus.

4   Conclusion

We detailed the eigenanatomy theory and algorithm and showed that eigenanatomy improves image reconstruction from a sparse set of anatomical basis functions. We showed that eigenanatomy also improves detection power for detecting group differences in longitudinal cortical change relative to SVD, PMD and univariate VBM. We note that showing that a method increases detection power does not comprise sufficient validation. This approach is not limited to longitudinal analysis and may be applied in a variety of morphometry contexts. Future extensions to the basic eigenanatomy algorithm given here will include network interpretations, exploration of alternatives to the SNLCG algorithm, alternative penalty terms and an automation of parameter selection.

References

1. Lee, D.D., Seung, H.S.: Learning the parts of objects by non-negative matrix factorization. Nature 401(6755), 788–791 (1999)
2. Zou, H., Hastie, T., Tibshirani, R.: Sparse principal component analysis. Journal of Computational and Graphical Statistics 15(2), 262–286 (2006)
3. Hoyer, P.O., Dayan, P.: Non-negative matrix factorization with sparseness constraints. Journal of Machine Learning Research 5, 1457–1469 (2004)
4. Guan, N., Tao, D., Luo, Z., Yuan, B.: Manifold regularized discriminative non-negative matrix factorization with fast gradient descent. IEEE Trans. Image Process. 20(7), 2030–2048 (2011)
5. Cai, D., He, X., Han, J., Huang, T.S.: Graph regularized non-negative matrix factorization for data representation. IEEE Trans. Pattern Anal. Mach. Intell. (December 2010)

6. Hosoda, K., Watanabe, M., Wersing, H., Krner, E., Tsujino, H., Tamura, H., Fujita, I.: A model for learning topographically organized parts-based representations of objects in visual cortex: topographic nonnegative matrix factorization. Neural Comput. 21(9), 2605–2633 (2009)

7. Witten, D.M., Tibshirani, R.: A framework for feature selection in clustering. J. Am. Stat. Assoc. 105(490), 713–726 (2010)

8. Friedman, J., Hastie, T., Tibshirani, R.: Regularization paths for generalized linear models via coordinate descent. J. Stat. Softw. 33(1), 1–22 (2010)

9. Cherkassky, V., Ma, Y.: Another look at statistical learning theory and regularization. Neural Netw. 22(7), 958–969 (2009)

10. Friedman, J., Hastie, T., Tibshirani, R.: Sparse inverse covariance estimation with the graphical lasso. Biostatistics 9(3), 432–441 (2008)

11. Yushkevich, P.A., Avants, B.B., Das, S.R., Pluta, J., Altinay, M., Craige, C., A.D.N.I.: Bias in estimation of hippocampal atrophy using deformation-based morphometry arises from asymmetric global normalization: an illustration in ADNI 3 t MRI data. Neuroimage 50(2), 434–445 (2010)

12. Holland, D., Dale, A.M., A.D.N.I.: Nonlinear registration of longitudinal images and measurement of change in regions of interest. Med. Image Anal. 15(4), 489–497 (2011)

13. 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)

14. Ashburner, J., Friston, K.J.: Unified segmentation. Neuroimage 26(3), 839–851 (2005)

15. Irwin, D.J., McMillan, C.T., Toledo, J.B., Arnold, S.A., Shaw, L.M., Wang, L.S., Trojanowski, J.Q., Lee, V.M.Y., Grossman, M.: Comparison of Cerbebrospinal Fluid Levels of Tau and ABeta1-42 in Alzheimer's Disease and Frontotemporal Degeneration Using Two Analytical Platforms. Archives of neurology (2011) (submitted)

16. Witten, D.M., Tibshirani, R., Hastie, T.: A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. Biostatistics 10(3), 515–534 (2009)

17. Kim, H., Park, H.: Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis. Bioinformatics 23, 1495–1502 (2007)

18. Heiler, M., Schnörr, C.: Learning sparse representations by non-negative matrix factorization and sequential cone programming. Journal of Machine Learning Research 7, 1385–1407 (2006)

19. d'Aspremont, A., El Ghaoui, L., Jordan, M.I., Lanckriet, G.R.G.: A direct formulation for sparse pca using semidefinite programming. SIAM Rev. 49, 434–448 (2007)

20. Lee, K., Tak, S., Ye, J.C.: A data-driven sparse glm for fMRI analysis using sparse dictionary learning with mdl criterion. IEEE Trans. Med. Imaging 30(5), 1076–1089 (2011)

21. Sill, M., Kaiser, S., Benner, A., Kopp-Schneider, A.: Robust biclustering by sparse singular value decomposition incorporating stability selection. Bioinformatics 27(15), 2089–2097 (2011)

Optimization of fMRI-Derived ROIs
Based on Coherent Functional Interaction Patterns

Fan Deng, Dajiang Zhu, and Tianming Liu

Department of Computer Science and Bioimaging Research Center,
The University of Georgia, Athens, GA

Abstract. Accurate localization of functionally meaningful Regions of Interests (ROIs) from fMRI data is critically important to functional brain imaging. A variety of established approaches such as General Linear Model (GLM) have been widely used in the community. How to determine the optimal location and size of an fMRI-derived ROI, however, remains an open, challenging problem. This paper presents a novel individualized optimization algorithm that simultaneously optimizes the locations and sizes of fMRI-derived ROIs by maximizing the coherences of their functional interaction patterns with respect to the block-based paradigm. As an alternative ROI optimization approach using functional interaction patterns, the algorithm was applied on a working memory task-based fMRI dataset and the experimental results are promising.

1   Introduction

Identifying functionally-specialized brain Regions of Interest (ROIs) constitutes the first step in a number of functional brain image analyses. For the purpose of functional ROI localization, task-based functional Magnetic Resonance Imaging (fMRI) has been widely considered as a benchmark approach in the community [1], e.g., via the General Linear Model (GLM) for activation detection [2]. Despite the wide use of this approach, there has been a significant, open question: how to pick the best possible fMRI signals to represent the activated region? In other words, given a roughly localized activation map, what are the optimal location and size of an fMRI-derived ROI?

Recently, a variety of studies demonstrated that the temporal functional connectivity curves on meaningful structural ROIs or landmarks roughly follow the external stimulus curve in task-based functional neuroimaging [3]-[5]. For instance, Lim et al. [3] has made an attempt in representing the functional brain states by concatenating functional connectivities on DTI-derived axonal fibers into feature vectors. Interestingly, via a sliding-window approach, they discovered that brain state changes in task-based fMRI are in temporal correspondence with the task paradigm [3]. In [4], Lindenberger et al. used electroencephalography (EEG) to measure the within-brain functional interaction. They found the interaction at the onset or starting point of certain events is strongest [4]. The authors in [5] reported that many functional connectivity curves on DTI-derived fibers closely follow the external stimulus task curves, and they can detect more activated brain regions than the raw fMRI BOLD signals.
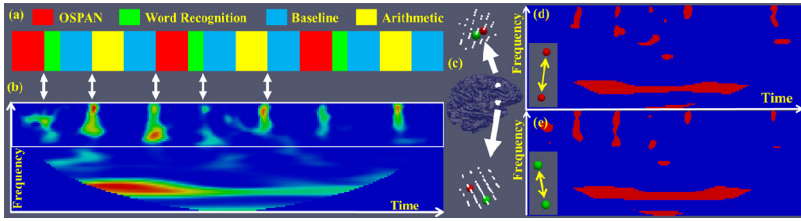
Fig. 1. (a) Working memory block-based paradigm. (b) Average interaction map between two ROIs. Red and blue represent strong and weak interaction, respectively. The block design is temporally aligned with the interaction map. (c)-(e): functional interaction maps from two pairs of time series selected from two ROIs in a working memory task-based session. (c) Location of the ROIs and the two pairs of voxels (pair 1 is in red and pair 2 is in green). (d) and (e): Functional interaction map of the pair 1 and 2 in (c), respectively.

Along the similar direction as the above studies, in our own experiments, we observed similar patterns in the time-frequency domain, as shown in Figs. 1a and 1b. Specifically, the working memory block-based paradigm and the average functional interaction map among ROIs (details in section 2.1) exhibited quite close temporal correlation, as shown in Figs. 1a and 1b, respectively. We can see that regions with intense interactions in the high frequency domain occur between adjacent blocks. This interesting and meaningful pattern has been replicated in other working memory block-based fMRI sessions we studied. As an example, Figs. 1c-1e show two ROIs identified by activation detection in a working memory paradigm. We randomly selected two pairs of time series from two ROIs (each pair contains one time series from each ROI), and depicted their functional interaction maps in Figs. 1d and 1e, respectively. Their functional interaction patterns are quite consistent. Therefore, we postulate that voxels in meaningful ROIs should possess fMRI BOLD signals revealing similar coherent functional interaction patterns with other voxels within this network.
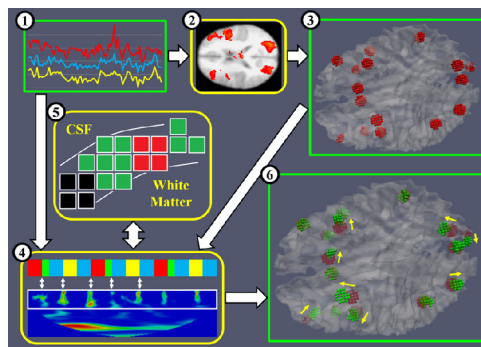

Fig. 2. Overview of the proposed algorithmic pipeline. (1) Task-based fMRI data. (2) GLM activation map. (3) Initial ROIs selected from activation peaks. (4) Functional coherence optimizer. (5) Spatial constraints. Red: current ROI; green: candidate voxels; dark: far away voxels are penalized. (6) Optimized ROIs (green). Red: initial positions; yellow arrows: movement.

Based on the abovementioned observation, we propose in this paper a novel ROI optimization algorithm that optimizes the locations and sizes of ROIs simultaneously, given the fMRI activation peaks to initialize. Our basic premise is the optimization should improve the coherence of meaningful functional interaction patterns among fMRI-derived ROIs. With the mathematical and computational modeling of the above principle, our proposed algorithm simultaneously optimizes the locations and sizes of ROIs by maximizing the coherences of functional interaction patterns with respect to the block-based paradigm.

2   Methods

An overview of the proposed algorithmic pipeline is shown in Fig. 2. Briefly, initial ROIs (3) are obtained from task-based fMRI data (1) through GLM (2); the optimizer builds a functional interaction coherence model (4) and estimates coherence values with spatial constraints (5) to simultaneously optimize the locations and sizes of ROIs (6). Within a single subject, all the ROIs are optimized at the same time.

2.1   Functional Interaction Model

We derived a functional interaction model from cross wavelet transform (XWT) [6], [7] to describe the time-frequency domain interactions. A coherence measure was then obtained for each pair of time series to evaluate the interaction patterns with respect to the block-based paradigm.

The core of the model, XWT, is a powerful tool in assessing the multi-scale time-frequency characteristics between time series. The basis functions, or wavelets, in wavelet transform, are localized both in the temporal and frequency domain, comparing with the sine wave basis functions in Fourier transform localized only in the frequency domain. The time-frequency result from a wavelet transform reveals the spectral characteristics at multiple frequencies or scales at different time of the time series. By combining the wavelet transforms, XWT uncovers the regions in the time-frequency domain where both time series share high co-power, which is considered and defined as the functional interaction in [6] and this paper. Furthermore, under a statistical significance test, XWT suggests the time and frequency/scale where the time series have intense interactions. Consider two time series of the same length, X and Y, their cross wavelet transform is defined as:

$$XWT_{XY} = W_X W_Y^*$$

(1)

where W denotes continuous wavelet transform (CWT); and W_Y^* denotes the complex conjugation of W_Y. We used the Morlet kernel [7] and statistical significance test of 95% significance level [6], [7]. With this configuration, we obtain a binary matrix representation for each pair of time series, denoted as an interaction map:

$$IM_{XY}(n,s) = \begin{cases} 1 & |XWT_{XY}(n,s)| \text{ is significant} \\ 0 & \text{otherwise} \end{cases}$$

(2)

where n is the time index and s is the scale index. Elements in IM_{XY} inside the Cone of Influence [6] are always set to zero due to the boundary effect of CWT. The binary representation enables us to focus on regions in the time-frequency domain where the co-power of the time series are significant. In fact, Fig. 1b is an averaged result of hundreds of interaction maps of each time series pair from two ROIs.

The high-frequency part of the interaction map is of special interest since it contains fine interaction patterns at various scales. We derive a binary histogram indicator array of interactions H_{XY} for the high-frequency part of the interaction map:

$$H_{XY}(n) = \begin{cases} 1 & \sum_{s>s_{lowerbound}} IM_{XY}(n,s) > 0 \\ 0 & otherwise \end{cases} \tag{3}$$

A one in the array indicates that the co-power of time series X and Y is significant at corresponding time and scale. A zero indicates no interaction in the interaction map.

The coherence measure was obtained by relating the binary indicator array with the block-based paradigm. For a paradigm with M blocks denoted as $\{B_1, B_2, ..., B_M\}$, there are M-1 transition states $\{S_1, S_2, ..., S_{M-1}\}$ with S_k being the transition state between the kth block and the $(k+1)^{th}$ block. A transition indicator array T_{XY} of size M-1 is derived from the histogram indicator array H_{XY} to represent the availability of any transition state in the interaction map. Specifically, $T_{XY}(k)=1$ if and only if $H_{XY}(n_1)=1$ and $H_{XY}(n_1)=1$, in which n_1 is the last time point of the kth block and n_1+1 is the first time point of the $(k+1)^{th}$ block. Simply put, there exist at least two adjacent ones in the binary histogram indicator array that covers the transition state k. We derive the coherence measure of time series X and Y as:

$$Q_{XY} = \frac{\sum_{k=1}^{M-1} T_{XY}(k)}{M-1} \tag{4}$$

Q_{XY} measures how coherent the observed interactions and the paradigm design are with regard to the state transitions. This pair-wise coherence measurement was extended to ROI-level and network-level through a linear combination.

$$Q_{ROI}(R_i, R_j) = \frac{1}{N_{R_i} N_{R_j}} \sum_{k_1=1}^{N_{R_i}} \sum_{k_2=1}^{N_{R_j}} Q_{R_i(k_1)R_j(k_2)} \tag{5}$$

where R_i and R_j are two ROIs; N_{R_i} and N_{R_j} are the number of voxels in R_i and R_j, respectively; $R_i(k_1)$ and $R_j(k_2)$ are the k_1^{th} and k_2^{th} voxel or time series in R_i and R_j, respectively. Q_{ROI} gives the ROI-level coherence measure. Meanwhile, the network-level coherence measure is given by:

$$Q = \frac{2}{N_{ROI}(N_{ROI}-1)} \sum_{1 \le i < j \le N_{ROI}} Q_{ROI}(R_i, R_j) \tag{6}$$

where N_{ROI} is the number of ROIs in the network. This network-level measure is essentially an average of the ROI-level measures of $N_{ROI}(N_{ROI}-1)/2$ possible ROI pairs.

Our primary focus in this optimization problem is to maximize the network-level coherence measure by simultaneously optimizing the location and size of each ROI.

2.2   Coherence Voting and Spatial Constraints

To assess the individual significance or coherence of every ROI voxel, we propagate the pair-wise coherence values to voxels using a voting strategy. The average vote is denoted as its individual coherence value. For the jth voxel in the ith ROI, its individual coherence is defined as:

$$Q_{ij} = \frac{1}{\sum_{\substack{1 \leq m \leq N_{ROI} \\ m \neq i}} N_{R_M}} \sum_{\substack{1 \leq k \leq N_{ROI} \\ k \neq i}} \sum_{l=1}^{N_{R_k}} Q_{R_i(j)R_k(l)} \tag{7}$$

This individual coherence value represents how coherent a voxel in an ROI interacts with all other ROIs in the network with regard to the block-design paradigm. Therefore it is a network-level measure of individual ROI voxels. By optimizing ROIs so that they include voxels with higher Q_{ij} values, the network-level coherence improves as well, since both Q_{ij} and Q are linear combinations of the same set of pair-wise coherence values with positive coefficients.

While the coherence measurement is meaningful to better localize the ROIs, it is possible that the variability and noise in fMRI data could still drag an ROI from its true location or change the size to include irrelevant voxels. A second concern regards our basic assumption: GLM already provides rough locations of the ROIs. A recent study showed that the fMRI-derived activation peak could be shifted a few voxels from its true location due to spatial smoothing [8]. Therefore, the optimization procedure should not move ROIs too far away from their initial locations. Hence we enforce three spatial constraints to regularize the optimization. (1) The search space of our optimization routine is limited to grey matter voxels to ensure we extract the right time series (grey matter constraint). (2) Voxels far away from the mass center of the current ROI are penalized to ensure the ROI does not grow too large (size constraint). (3) Voxels far away from the initial location are penalized (movement constraint).

2.3   ROI Optimization

Our goal in this optimization problem is to move and/or reshape ROIs to include voxels with higher individual coherence values with spatial constraints. A unified procedure was employed to add qualified voxels into a ROI or removes certain voxels from a ROI. For a candidate voxel v for ROI R_i, its individual coherence value is converted to an intrinsic weight:

$$p_{vote}(v) = \exp(-2\left\|\overline{Q}_{iv} - 1\right\|^2 \big/ \sigma_{R_i}) \tag{8}$$

where σ_{Ri} is the standard deviation of all individual coherence values in ROI R_i; $\overline{Q}_{iv}$ is the normalized individual coherence value of the candidate voxel v with regard to ROI R_i, $0 \leq \overline{Q}_{iv} \leq 1$.

Finally, the individual coherence value and the spatial constraints are combined for each candidate voxel. We obtain the following probability of removal:

$$p_{removal}(v) = p_{movement}(v)\, p_{size}(v)\, p_{vote}(v) \tag{9}$$

where p_{size} and $p_{movement}$ are ROI size penalty and movement penalty, respectively.

Our iterative optimization process is as follows.

Initialization. Grey matter voxels within 3-voxel distance to the initial locations from GLM are included. Note the results are insensitive to the initial size.

Optimization. In each iteration, only voxels at the surface of each ROI are changed since we do not want to create topological issues, e.g. a hole inside a ROI. The actual optimization is a two-step procedure. In the first step, all grey matter voxels within the neighborhood of surface voxels with probabilities of removal smaller than a preselected threshold are included in the optimized ROI. In our experiments, we set the threshold to 0.3. The probabilities and weighted mass centers are recalculated. In the second step, all surface voxels with probabilities of removal larger than a preselected threshold are removed. In our experiments, we set the threshold to 0.7. The optimization procedure ends when the ROIs have been stabilized. Our algorithm processes all ROIs simultaneously and searches within the whole space (activated regions). It is a global search that ensures optimal, instead of suboptimal, solution.
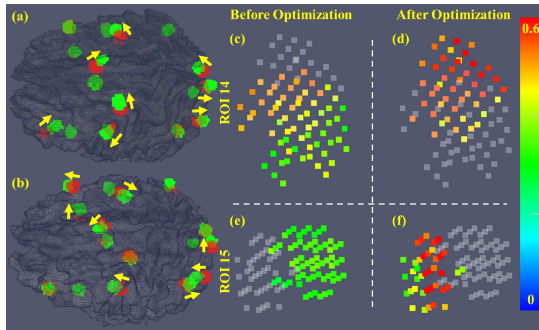

Fig. 3. (a) and (b): Initial ROIs (red) and optimized ROIs (green) of two subjects. Yellow arrows indicate rough directions of movement. The average movement is 3.50 mm. (c)-(f): individual coherence measures of two ROIs before and after optimization. (c) and (e): ROI 14 and 15 before optimization. (d) and (f): ROI 14 and 15 after optimization. The voxels are colorcoded in their individual coherence measure values. The voxels in transparency are from corresponding optimized ROIs in (c) and (e); and are from corresponding initial ROIs in (d) and (f) for easier visual comparison.

3   Results

We applied the proposed algorithm on a working memory fMRI dataset with twelve healthy subjects. The multimodal fMRI and DTI data (only for validation) are analyzed by FSL [8]-[10]. Sixteen initial ROIs were derived from GLM results.

3.1   Optimized ROIs

Examples of optimized ROIs are shown in Figs. 3a and 3b. We can see that the optimized ROIs are quite close to their initial locations. The overall average movement in twelve subjects is 3.50mm (standard deviation: 2.54mm). The average size of the optimized ROIs is 40.16 voxels (standard deviation: 6.53 voxels), a 29.52% drop comparing with initial ROIs. The network-level coherence was increased by 14.87% on average, proving that our algorithm is effective in moving and reshaping the ROIs. Shown in Figs. 3c-3f are examples of individual coherence measures. We can see that the optimized ROIs contain voxels with higher coherence values in comparison with the initial ROIs. This visualization further proves the effectiveness of our algorithm.

3.2   Signal Consistency

Average Z values from the GLM activation detection were calculated for each subject. Overall, we did not found any significant difference after optimization. The comparable Z values before and after optimization confirms that the optimized ROIs are still in the activated regions. We also computed the ratio of the variance of the first (principal) component using Principal Component Analysis (PCA). The average ratio before and after optimization are 98.3% and 98.2%, respectively. The overall functional connectivity measured by Pearson correlation increased by 0.02. These results are in line with our expectation to the optimization, which is, increasing the functional coherence while maintaining the signal consistency.
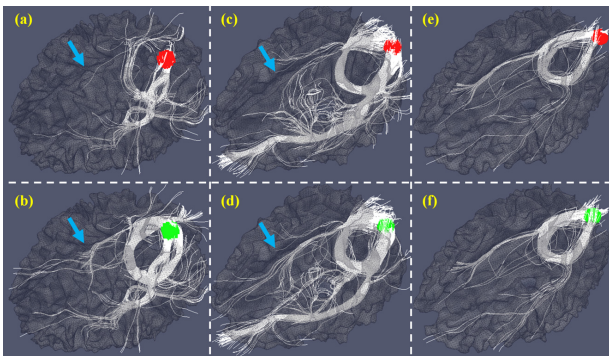

Fig. 4. Visualization of fibers connecting to the same ROI before and after optimization for three subjects. Each column is one subject. (a), (c), and (e): before optimization; (b), (d), and (f): after optimization. Red: initial ROI; green: optimized results.

3.3   Structural Connectivity

As an independent validation study, we used structural connectivity pattern derived from DTI data to measure the improvements. Specifically, we implemented a similar algorithm as the trace map model [11] to quantitatively measure the fiber connections.

For fair comparisons, we used FSL FLIRT [12] to register the extracted fibers to a template subject. For every pair of subjects we calculated for each ROI the distance between their trace map features. The average distance for each ROI was recorded. It is interesting that our optimization based on coherent functional interaction patterns indeed improves the overall structural consistency measured by the trace-map model, by 4.26%. As an example, we visualized the fibers connecting to one ROI for three subjects in Fig. 4. It is evident that after optimization, the fibers emanating from the same ROI become more consistent (cyan arrows). This further confirms the validity of our optimization: both functional and structural connectivity improve.

4   Conclusion

In this paper we proposed to simultaneously optimize the locations and sizes of GLM-derived ROIs based on functional interactions between fMRI time series. Apart from the traditional GLM-based approach, a novel wavelet-based interaction model was designed to capture the intrinsic interactions and make full use of the block-based paradigm by relating the block-based paradigm with high-frequency functional interactions. The optimization algorithm is evaluated and validated using GLM activation detection results, functional coherence metric, functional connectivity consistency, and structural connectivity consistency, and promising results were achieved.

With the availability of these optimized functional ROIs, we will be able to extract more accurate and reliable fMRI signals from these ROIs to enable and facilitate other important tasks such as assessment of functional dynamics and interactions among brain networks and construction of predictive models of functional brain ROIs based on consistent structural fiber connection patterns. Our current algorithm is individualized. group-wise treatment of those ROIs requires huge amount of combinational processing and optimization across multiple brains and is a big challenge at the current stage. In addition, it is very difficult to compare fMRI signals across individuals due to the vast complexity and variability. In the future, we plan to further investigate into the group-wise extension of the proposed approach. Our future work also includes applying this approach on more fMRI paradigms and jointly optimizing ROIs with structural constraints.

References

1. Poldrack, R.A.: Region of interest analysis for fMRI. Soc. Cogn. Affect. Neurosci. 2(1), 67–70 (2007)
2. Friston, K.J.: Modalities, modes, and models in functional neuroimaging. Science 326(5951), 399–403 (2009)
3. Lim, C., et al.: Brain state change detection via fiber-centered functional connectivity analysis. In: ISBI (2011)
4. Lindenberger, U., et al.: Brains swinging in concert: cortical phase synchronization while playing guitar. BMC Neurosci. 10(22) (2009)

5. Lv, J., Guo, L., Li, K., Hu, X., Zhu, D., Han, J., Liu, T.: Activated Fibers: Fiber-Centered Activation Detection in Task-Based FMRI. In: Székely, G., Hahn, H.K. (eds.) IPMI 2011. LNCS, vol. 6801, pp. 574–587. Springer, Heidelberg (2011)
6. Torrence, C., Compo, G.P.: A practical guide to wavelet analysis. B. Am. Meteorol. Soc. 79(1), 61–78 (1998)
7. Chang, C., Glover, G.H.: Time-frequency dynamics of resting-state brain connectivity measured with fMRI. NeuroImage 50(1), 80–98 (2010)
8. Anonymous
9. Woolrich, M.W., Ripley, B.D., Brady, J.M., Smith, S.M.: Temporal autocorrelation in univariate linear modelling of FMRI data. NeuroImage 14(6), 1370–1386 (2001)
10. Anonymous
11. Zhu, D., et al.: Optimization of functional brain ROIs via maximization of consistency of structural connectivity profiles. NeuroImage (2011)
12. Jenkinson, M., Smith, S.M.: A global optimisation method for robust affine registration of brain images. Med. Image Anal. 5(2), 143–156 (2001)

Topology Preserving Atlas Construction from Shape Data without Correspondence Using Sparse Parameters

Stanley Durrleman[1], Marcel Prastawa[2], Julie R. Korenberg[3], Sarang Joshi[2], Alain Trouvé[4], and Guido Gerig[2,3]

[1] INRIA / ICM, Pitié Salpêtrière Hospital, Paris, France
[2] SCI Institute, University of Utah, Salt Lake City, USA
[3] Brain Institute, University of Utah, Salt Lake City, USA
[4] CMLA - Ecole Normale Supérieure de Cachan, Cachan, France

Abstract. Statistical analysis of shapes, performed by constructing an atlas composed of an average model of shapes within a population and associated deformation maps, is a fundamental aspect of medical imaging studies. Usual methods for constructing a shape atlas require point correspondences across subjects, which are difficult in practice. By contrast, methods based on currents do not require correspondence. However, existing atlas construction methods using currents suffer from two limitations. First, the template current is not in the form of a topologically correct mesh, which makes direct analysis on shapes difficult. Second, the deformations are parametrized by vectors at the same location as the normals of the template current which often provides a parametrization that is more dense than required. In this paper, we propose a novel method for constructing shape atlases using currents where topology of the template is preserved and deformation parameters are optimized independently of the shape parameters. We use an L^1-type prior that enables us to adaptively compute sparse and low dimensional parameterization of deformations. We show an application of our method for comparing anatomical shapes of patients with Down's syndrome and healthy controls, where the sparse parametrization of diffeomorphisms decreases the parameter dimension by one order of magnitude.

1 Introduction

Shape statistics is fundamental to medical imaging studies, for instance to characterize normal versus pathological structures or to highlight the effect of treatments on anatomical structures. Usual methods build an average shape model, called template, that is representative of a shape ensemble, and correspondences between the template and each subject's shape, altogether called an atlas. Differences appear in the shape model used and the way shapes are put into correspondence. Parametric models, such as medial axis representation [12], extract low-dimensional shape features, which are averaged and compared across subjects. Non parametric methods are often based on point correspondences across

shapes with an either fixed or optimized parameterization of the curves or surfaces [2,7,9,13]. Homologous point positions are averaged and 3D-deformations map the average model to the point configuration of each subject, such maps being used for deriving shape statistics [14,11]. To alleviate the problem of finding point correspondences across shapes, which is not always possible in practice, one can compare shapes using the metric on currents [15]. In this framework, shape statistics are represented by a template current and diffeomorphic 3D mappings that map the template current to each subject's shape.

Existing atlas construction approaches using the metric on currents [6,4,3] suffer from two main limitations. First, the template current is either the superimposition of warped surfaces [6] or a set of disconnected normals or tangents (Dirac currents) [4,3] and therefore is not given as a mesh. It has never been shown that such templates could be the approximation of a continuous surface. In this sense, these approaches do not preserve the topology of the shapes, namely the continuity of the surfaces and the number of connected components of the shapes in the population, thus limiting its practical use for segmentation purposes, for instance. Second, the template-to-surface deformations are parameterized by momentum or speed vectors that are at the same location as the normal or tangent of the template current. However, there is no reason to link the parameterization of the template shape with the parameterization of the deformations. The most variable parts of the shape are not necessarily the parts that require the finest sampling, e.g. the most folded ones. Therefore, the template shape and the deformations should have independent parameterization.

In this paper, we propose to build a shape atlas using the metric on currents, but in a way which allow the user to fix the topology of the template mesh, namely the number of vertices and the edges connecting them, and to optimize their position in the 3D space. This will be possible due to an optimal control formulation of the diffeomorphic matching problem in the Large Deformation Diffeomorphic Metric Mapping paradigm [10,5]. In this formulation, the three variables to optimize, namely the template shape, the control points positions and the momentum vectors attached to them that parameterize the deformations, are seen as the initial conditions of a dynamical system. We derive here the analytic expression of the gradient with respect to these variables. Additionally, we use a L^1 prior to select the most relevant subset of control points for the parameterization of the shape variability. Such a sparse and adaptive parameterization will be particularly well suited for statistical purposes. We show an application of our method to a study comparing anatomical features of patients with Down's syndrome (DS) and healthy controls, where the sparse parametrization of shape variability enable statistical analysis in lower dimensional spaces.

2   Shape Atlas Construction

2.1   Joint Estimation of Template and Deformations

The method aims to estimate a template shape $\mathbf{X}_0$ and template-to-subjects deformations ϕ_i from a set of shapes $\mathbf{X}_1, \ldots, \mathbf{X}_{N_{\mathrm{su}}}$. Each $\mathbf{X}_i$ denotes a vector

containing the positions of the vertices of the input shapes, which may be of different sizes. $\mathbf{X}_0$ denotes the vertices of the template shape, whose number and connectivity is fixed by the user. Only the positions of the vertices of the template are optimized and not its connectivity, so that the topology of the template is preserved during optimization. In the framework of Fréchet means [11], we estimate one template and N_{su} deformations ϕ_i that minimize the criterion:

$$E(\mathbf{X}_0, \{\phi_1, \ldots, \phi_{N_{\mathrm{su}}}\}) = \sum_{i=1}^{N_{\mathrm{su}}} D(\phi_i(\mathbf{X}_0), \mathbf{X}_i) + \mathrm{Reg}(\phi_i), \qquad (1)$$

where D denotes the squared distance on currents, $\phi_i(\mathbf{X}_0)$ the deformation of the template shape $\mathbf{X}_0$ and Reg a measure of regularity of the deformations.

2.2   Parameterization of Deformations

We use a mechanical system of self-interacting particles, called control points, to build dense diffeomorphic deformations [5]. Let $\mathbf{c}_0 = \{c_{0,k}\}$ be a set of control points and $\boldsymbol{\alpha}_0 = \{\alpha_{0,k}\}$ a set of initial momenta of the particles, altogether called the initial state of the system $\mathbf{S}_0 = \{\mathbf{c}_0, \boldsymbol{\alpha}_0\}$. This set of particles moves from time $t = 0$ to $t = 1$ according to following equations of motion:

$$\begin{cases} \dot{c}_k(t) = \sum_{p=1}^{N} K(c_k(t), c_p(t))\alpha_p(t) \\ \dot{\alpha}_k(t) = -\sum_{p=1}^{N} \alpha_k(t)^t \alpha_p(t) \nabla_1 K(c_k(t), c_p(t)) \end{cases}, \qquad (2)$$

which are such that the energy of the system $\sum_{i,j} \alpha_i(t)^t K(c_i(t), c_j(t))\alpha_j(t)$ is conserved in time. The kernel K models the interaction forces among the particles. These equations describe the evolution of the state of the system $\mathbf{S}(t) = \{c_k(t), \alpha_k(t)\}$ and can be written in short: $\dot{\mathbf{S}}(t) = F(\mathbf{S}(t))$ with the initial condition $\mathbf{S}(0) = \mathbf{S}_0$.

The motion of the control points defines a diffeomorphism of the whole 3D space [8]. The speed at position x_0 interpolates the speed of the control points:

$$\dot{x}(t) = \sum_{k=1}^{N} K(x(t), c_k(t))\alpha_k(t). \qquad (3)$$

This equation shows that the rate of decay of the kernel determines the size of the neighborhood that is "pulled" by each control point. It can be written in short as $\dot{x}(t) = G(x(t), \mathbf{S}(t))$ with the initial condition $x(0) = x_0$. Using the vertices of the template shape x_0 as initial conditions, the integration of this equation computes the deformation of the template shape $\mathbf{X}_0 = \mathbf{X}(0)$ to $\mathbf{X}(1)$ that is equal to $\phi(\mathbf{X}_0)$ with ϕ is the diffeomorphism parameterized by $\mathbf{S}_0$.

2.3   Atlas Estimation

The N_{su} template-to-subject deformations in the criterion (1) are parameterized by N_{su} vectors $\mathbf{S}_0^i$. Each of these vectors serves as the initial condition in (2). Then, the template deformation is obtained by integration of (2) followed by (3). We choose to use the same set of control points $\mathbf{c}_0$ for all subjects, which defines a common basis for the deformations parameterization. By contrast, the set of initial momenta $\boldsymbol{\alpha}_0^i$ are subject-specific. As a regularizer of the deformations, we use the energy of the set of particles $\sum_{p,q} \alpha_{0,p}^i{}^t K(c_{0,p}, c_{0,q})\alpha_{0,q}^i$, which is also the geodesic distance connecting the template to the the ith shape [8]. In order to select the most relevant subset of control points, we add a L^1 penalty to (1), so that the criterion to minimize writes:

$$E(\mathbf{X}_0, \mathbf{c}_0, \{\boldsymbol{\alpha}_0^i\}) = \sum_{i=1}^{N_{\text{su}}} \left\{ \frac{1}{2\sigma^2} D(\mathbf{X}_0^i(1), \mathbf{X}_i) + \sum_{p,q} \alpha_{0,p}^i{}^t K(c_{0,p}, c_{0,q})\alpha_{0,q}^i + \gamma \sum_p \left\| \alpha_{0,p}^i \right\| \right\}$$

(4)

subject to:

$$\begin{cases} \dot{\mathbf{S}}^i(t) = F(\mathbf{S}^i(t)) & \text{with } \mathbf{S}^i(0) = \{\mathbf{c}_0, \boldsymbol{\alpha}_0^i\} \\ \dot{\mathbf{X}}_0^i(t) = G(\mathbf{X}_0^i(t), \mathbf{S}^i(t)) & \text{with } \mathbf{X}_0^i(0) = \mathbf{X}_0 \end{cases}$$

(5)

The first equation in (5) is the equations of motion of the particles, like in (2). The second equation is deformation of the template parameterized by the particles motion, like in (3). σ^2 and γ balance the data term against the regularization terms. The variables to be optimized are: (i) the position of the vertices of the template shape $\mathbf{X}_0$, (ii) the position of the control points in the template domain $\mathbf{c}_0$ and (iii) the N_{su} initial momenta $\boldsymbol{\alpha}_0^i$ that parameterize each template-to-subject deformation. In practice, we also regularize the template shape defined by $\mathbf{X}_0$ by applying a penalty on Gaussian curvature of the mesh.

Only the first two terms in (4) are differentiable. As shown in the supplementary material accessible at the first author's webpage, the gradient of data term is given as:

$$\nabla_{\boldsymbol{\alpha}_0^i} D = \xi^{\alpha,i}(0) \quad \nabla_{\mathbf{c}} D = \sum_{i=1}^{N_{\text{su}}} \xi^{c,i}(0) \quad \nabla_{\mathbf{X}_0} D = \sum_{i=1}^{N_{\text{su}}} \theta^i(0)$$

where the auxiliary variables $\xi^i(t) = \{\xi^{c,i}(t), \xi^{\alpha,i}(t)\}$ (of the same size as $\mathbf{S}^i(t)$) and $\theta^i(t)$ (of the same size as $\mathbf{X}_0$) satisfy the linear ODEs:

$$\dot{\theta}^i(t) = -(\partial_1 G(\mathbf{X}_0^i(t), \mathbf{S}^i(t)))^t \theta^i(t) \text{ with } \theta^i(1) = \frac{1}{2\sigma^2} \nabla_{\mathbf{X}_0^i(1)} D(\mathbf{X}_0^i(1), \mathbf{X}_i)$$

$$\dot{\xi}^i(t) = -\left(\partial_2 G(\mathbf{X}_0^i(t), \mathbf{S}^i(t))^t \theta^i(t) + d_{\mathbf{S}^i(t)} F(\mathbf{S}^i(t))^t \xi^i(t) \right) \text{ with } \xi^i(1) = 0$$

(6)

To compute the gradient, one integrates the flow equations (5) forward in time to build the deformations of the template shape. Then, one computes the gradient of the residual data term ∇D, which serves as initial conditions in (6).

The ODEs (6) transport this information from each subject's space back to the template space, where the final value of the auxiliary variables $\theta(t)$ (resp. $\xi(t)$) is used to update the template (resp. the control point positions and momenta).

To optimize (4), which combines differentiable terms denoted E_{L^2} with an L^1 penalty, we use an adapted gradient-descent scheme called Fast Iterative Shrinkage and Thresholding Algorithm [1]. The template and control points are not affected by the L^1 term and are updated using a gradient-descent step at each iteration. By contrast, the momenta $\alpha_{0,p}^i$ are updated according to:

$$\alpha_{0,p}^i \leftarrow S_{\tau\gamma}\left(\left\|\alpha_{0,k}^i - \tau\nabla_{\alpha_{0,k}^i}E_{L^2}\right\|\right)\frac{\alpha_{0,k}^i - \tau\nabla_{\alpha_{0,k}^i}E_{L^2}}{\left\|\alpha_{0,k}^i - \tau\nabla_{\alpha_{0,k}^i}E_{L^2}\right\|}, \tag{7}$$

where τ is the current step-size of the gradient descent and S the usual soft-thresholding function $S_\lambda(x) = \max(0, x - \lambda) + \min(0, x + \lambda)$. This function zeroes out the momenta that are too small in magnitude, thus ensuring sparsity in the parametrization of deformations.

The parameters of the algorithm are the trade-offs σ and γ, the standard deviation of the Gaussian kernels for the momenta σ_V and the currents metric σ_W [15,4]. We initialize the template shape with an ellipsoid for each connected component of the shapes, the control points with a regular lattice of step σ_V, and the initial momenta are set to zero.

3   Results

We apply our method to a study that seeks to compare neuroimaging, genetics, and neurotransmitter properties of patients with Down's syndrome and healthy controls. We construct an atlas from surfaces of three different deep brain structures: amygdala, hippocampus, and putamen (Fig. 1). We initialize the atlas with one ellipsoid for each of the three anatomical structures, and we initialize the control points with a regular lattice of 650 points. After the optimization, the template shapes capture the common anatomical features across the populations (Fig. 2-left), and are given as meshes with the same topology as the initial set of ellipsoids. The parameterization of the template-to-subject deformations were also optimized: control points are moved toward the surfaces (where their theoretical optimal locations are, as shown in [15]) and the sparsity prior selects a subset of control points (99 out of 650) that carry a non-zero momentum vector. Results were generated using $\sigma = 10^{-4}$, $\gamma = 3 \times 10^6$, $\sigma_V = 5$, and $\sigma_W = 2$.

By comparison, we show the atlas built with the method of [4] in Fig. 3. The template shape is given as a set of disconnected normals: the continuity of the surfaces and the number of connected components of the input shapes have not been preserved. Moreover, the template-to-subjects deformations are parameterized with momenta located at the same place as the template normals. By contrast, our method optimized the position and the number of the control
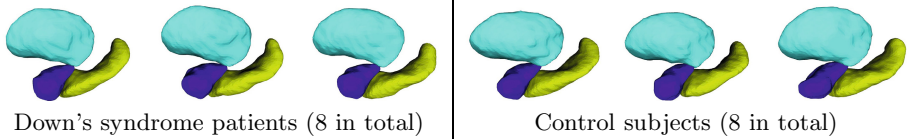
Fig. 1. Sample input shapes where hippocampus, amygdala, and putamen are shown in yellow, blue, and cyan respectively
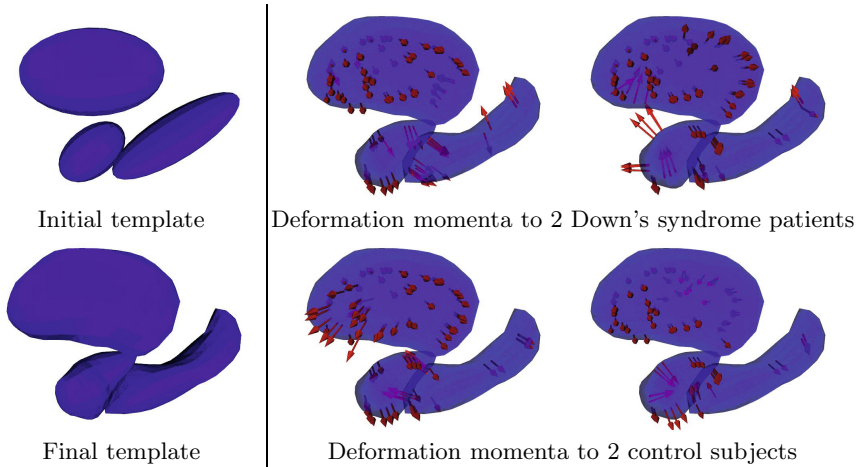

Fig. 2. Atlas construction: template shape (left) and parameterization of the template-to-subject deformations (right). The template is initialized with an ellipsoid per connected components (top-left). After the optimization, the template shapes still have the same topology as the sample shapes. Simultaneously, template-to-subjects deformations are estimated, which are parameterized by the momentum vectors (red arrows). The vectors are located at the position of the control points, which are the same for all subjects. Control point positions were initialized as the nodes of a regular lattice, and our algorithm moves the control points to their optimal position near the surfaces and selects the most relevant ones according to the sparsity prior.

points independently of the vertices of the template shapes. Consequently, deformations are parameterized by more than 40 times fewer momenta. Constraining the template to remain a mesh has not introduced bias in the estimation: the norm of the difference between the templates generated by the two methods is 3.4×10^{-5}, which is much smaller than the standard deviation and below the usual threshold of 3 times standard deviation to decide statistical significance.

We construct a common template for the combined DS and healthy populations, and perform a Principal Component Analysis (PCA) on the momentum vectors for each population separately. The results in Fig. 4 show that the two populations contain different variability at different objects and at different locations within each object. The sparse momentum vectors enable statistical analysis in lower dimensional space which has great potential for clinical studies.
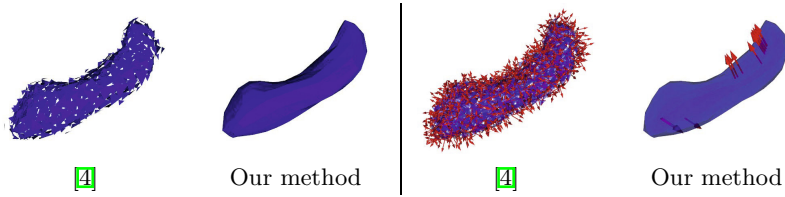
Fig. 3. Comparison of hippocampus template generated using [4] and our method. Arrows indicate the momenta driving the registration of the template to the first subject. Our method generates topologically correct mesh as opposed to triangle normals, and with momenta that are not constructed to be located at the surface triangles.
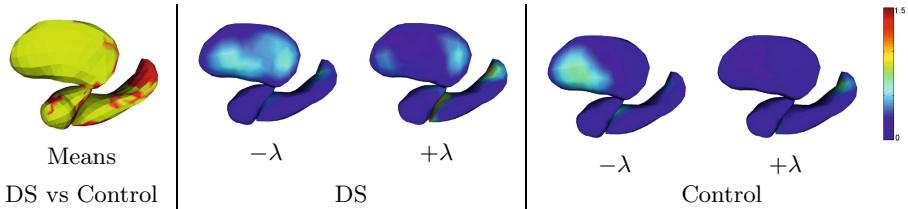

Fig. 4. Statistical analysis of momentum vectors for Down's syndrome (DS) and healthy control population. Left: template deformed according to the average momenta for DS (red) and control (yellow). Center and Right: Templates deformed according to the first mode of PCA on momentum vectors for DS and control population, demonstrating the different modes of variability. Colors indicate the magnitude of displacement from the population mean shape (in mm).

4   Conclusions

We propose a new method for estimating shape atlases using the currents metric. In contrast to [6,4], the template shape is given as a mesh that has the same topology as the shapes in the population and the parameterization of the template-to-subject deformations were optimized independently of the template shape. We use a single gradient descent for estimating the template and the deformations, in contrast to the alternating minimization in [6,4]. This scheme is much more efficient than using the matching pursuit technique and shows good convergence properties, even with the most naive initialization. Our non-parametric method makes use of the metric on currents, which enables the use of shapes without point correspondences, and therefore with minimal pre-processing.

The method also provides automatic parametrization of diffeomorphisms for mapping subjects within a population, where the parameters are constrained to lower dimensional spaces. We demonstrated the potential for these sparse parametrizations for performing statistical analysis on shapes from Down's syndrome and healthy control population groups, which can enable future research correlating brain function, anatomy, and neurocircuitry.

In the future, we plan to analyze the robustness and statistical power of the parametrizations provided by our method, in particular in High Dimension Low Sample Size (HDLSS) settings that are typical in imaging studies of populations.

References

1. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM Journal on Imaging Sciences 2(1), 183–202 (2009)
2. Chui, H., Rangarajan, A., Zhang, J., Leonard, C.: Unsupervised learning of an atlas from unlabeled point-sets. IEEE PAMI 26(2), 160–172 (2004)
3. Durrleman, S., Fillard, P., Pennec, X., Trouvé, A., Ayache, N.: Registration, atlas estimation and variability analysis of white matter fiber bundles modeled as currents. NeuroImage 55(3), 1073–1090 (2011)
4. Durrleman, S., Pennec, X., Trouvé, A., Ayache, N.: Statistical models of sets of curves and surfaces based on currents. Med. Image Anal. 13(5), 793–808 (2009)
5. Durrleman, S., Prastawa, M., Gerig, G., Joshi, S.: Optimal Data-Driven Sparse Parameterization of Diffeomorphisms for Population Analysis. In: Székely, G., Hahn, H.K. (eds.) IPMI 2011. LNCS, vol. 6801, pp. 123–134. Springer, Heidelberg (2011)
6. Glaunès, J., Joshi, S.: Template estimation from unlabeled point set data and surfaces for computational anatomy. In: Proc. of the International Workshop on the Mathematical Foundations of Computational Anatomy, MFCA 2006 (2006)
7. Hufnagel, H., Pennec, X., Ehrhardt, J., Ayache, N., Handels, H.: Generation of a statistical shape model with probabilistic point correspondences and EM-ICP. International Journal of CARS (IJCARS) 2(5), 265–273 (2008)
8. Joshi, S., Miller, M.: Landmark matching via large deformation diffeomorphisms. IEEE Trans. Image Processing 9(8), 1357–1370 (2000)
9. Kurtek, S., Klassen, E., Ding, Z., Jacobson, S., Jacobson, J., Avison, M., Srivastava, A.: Parameterization-invariant shape comparisons of anatomical surfaces. Trans. Med. Imag. 30, 849–858 (2011)
10. Marsland, S., McLachlan, R.: A Hamiltonian Particle Method for Diffeomorphic Image Registration. In: Karssemeijer, N., Lelieveldt, B. (eds.) IPMI 2007. LNCS, vol. 4584, pp. 396–407. Springer, Heidelberg (2007)
11. Pennec, X., Fillard, P., Ayache, N.: A Riemannian framework for tensor computing. International Journal of Computer Vision 66(1), 41–66 (2006)
12. Pizer, S., Fletcher, P., Joshi, S., Thall, A., Chen, J., Fridman, Y., Fritsch, D., Gash, A., Glotzer, J., Jiroutek, M., Lu, C., Muller, K., Tracton, G., Yushkevich, P., Chaney, E.: Deformable M-Reps for 3D Medical Image Segmentation. International Journal of Computer Vision 55(2-3), 85–106 (2003)
13. Székely, G., Kelemen, A., Brechbühler, C., Gerig, G.: Segmentation of 2-D and 3-D objects from MRI volume data using constrained elastic deformations of flexible fourier contour and surface models. Medical Image Analysis 1, 19–34 (1996)
14. Vaillant, M., Miller, M., Younes, L., Trouvé, A.: Statistics on diffeomorphisms via tangent space representations. NeuroImage 23, 161–169 (2004)
15. Vaillant, M., Glaunès, J.: Surface Matching via Currents. In: Christensen, G.E., Sonka, M. (eds.) IPMI 2005. LNCS, vol. 3565, pp. 381–392. Springer, Heidelberg (2005)

Dominant Component Analysis
of Electrophysiological Connectivity Networks

Yasser Ghanbari[1], Luke Bloy[1], Kayhan Batmanghelich[1],
Timothy P.L. Roberts[2,*], and Ragini Verma[1]

[1] Section of Biomedical Image Analysis, University of Pennsylvania, Philadelphia, PA
{Yasser.Ghanbari,Luke.Bloy,Nematollah.Batmanghelich,
Ragini.Verma}@uphs.upenn.edu
[2] Lurie Family Foundation's MEG Imaging Center, Department of Radiology,
Children's Hospital of Philadelphia, Philadelphia, PA, USA
robertstim@email.chop.edu

Abstract. Connectivity matrices obtained from various modalities (DTI, MEG and fMRI) provide a unique insight into brain processes. Their high dimensionality necessitates the development of methods for population-based statistics, in the face of small sample sizes. In this paper, we present such a method applicable to functional connectivity networks, based on identifying the basis of dominant connectivity components that characterize the patterns of brain pathology and population variation. Projection of individual connectivity matrices into this basis allows for dimensionality reduction, facilitating subsequent statistical analysis. We find dominant components for a collection of connectivity matrices by using the projective non-negative component analysis technique which ensures that the components have non-negative elements and are non-negatively combined to obtain individual subject networks, facilitating interpretation. We demonstrate the feasibility of our novel framework by applying it to simulated connectivity matrices as well as to a clinical study using connectivity matrices derived from resting state magnetoencephalography (MEG) data in a population of subjects diagnosed with autism spectrum disorder (ASD).

1 Introduction

Although the neurological origin of many brain disorders is still unknown, computational techniques applied to neuroimaging data help unveil the underlying functional or structural differences between patient and typically developing (TD) populations. Many brain disorders such as ASD and schizophrenia are believed to be due, in part, to altered brain connectivity [1–4]. Hence, connectivity analyses of brain function has received considerable attention as an indicator of or biomarker for brain disorders. Such studies have investigated the functional or structural connectivity networks using fMRI, DTI, EEG, or MEG recordings

by calculating correlation, synchronization, or mutual information measures and interpreting abnormalities such as deficient long-range connections [1, 3].

A number of analysis approaches exists to utilize connectivity methods to investigate the underlying brain processes occurring during rest as well as under specific task conditions. A successful analysis methodology must possess a means of identifying relevant sub-networks providing an interpretable representation of the brain activity, while also facilitating the statistical analysis able to describe how this representation is affected by disease. The approach taken here is the decomposition of connectivity matrices into dominant components while enforcing positivity of both the components and coefficients. Such a decomposition maintains the interpretation of each component as a connectivity matrix and the coefficients as activations of those networks while providing a succinct low dimensional representation of the population amenable to statistical analysis. The traditional approaches, principal and independent components analyses (PCA and ICA), used for investigating brain networks [5, 6] provide dimensionality reduction but lack the necessary constraints that provide a physiologic interpretability of the decomposition.

In this paper, we present the projective non-negative component analysis to identify the dominant functional connectivity networks in a population. These networks form a basis of variation in the population that could be due to differences in age, pathology, gender, etc. The non-negativity constraint on both components and the coefficients allows the interpretation of a difference in the coefficient of a particular component as a change in the degree to which it is activated. Additionally, the connectivity components are nearly orthogonal. The orthogonality of components with non-negative elements happens only when they do not share non-zero dimensions, leading to sparsity [7, 8]. Such unique analysis enables us to obtain the dominant networks, thereby providing a global view of the brain processes and how they are affected by disease.

The applicability of the proposed method is demonstrated in simulated connectivity matrices while comparing it with an alternate form of basis decomposition. While the method is generalizable to any type of connectivity matrix from alternate modalities, in this work we apply it to electrophysiological functional connectivity networks computed from resting-state MEG data by using the synchronization likelihood (SL) analysis [9]. This novel approach utilizes the high temporal resolution of MEG data, allowing an accurate characterization of the functional coupling between different areas of the brain. When coupled with a suitable analysis framework, such as the proposed, MEG based connectivity promises valuable insights into the temporal dynamics of brain states that is unavailable from other modalities as well as how they are affected by pathology.

2   Methods

2.1   Projective Non-Negative Component Analysis (PNCA)

We hypothesize that each connectivity matrix obtained from recording the brain activity of a subject is a linear combination of several fundamental connectivity

matrices called connectivity components. Due to the symmetry of connectivity matrices, a vector of all elements of the upper triangular part of any connectivity matrix is considered as a representative of that matrix, and is used as an observation vector $\boldsymbol{y}_i$ for the corresponding subject i. To compute the connectivity components whose mixture approximately constructs the observed connectivity matrices, a matrix factorization model is used as follows.

$$Y \approx W\boldsymbol{\Phi}, \tag{1}$$

where columns of $\boldsymbol{Y}_{p \times q}$, i.e. $\boldsymbol{y}_i$ $(1 \leq i \leq q)$, are the SL connectivity matrix representatives, and columns of $\boldsymbol{W}_{p \times r}$, i.e. $\boldsymbol{w}_j$ $(1 \leq j \leq r)$, are representative of the basis connectivity components, i.e. the upper triangular elements of the matrix of the corresponding connectivity component. These components $(\boldsymbol{w}_j)$ are then mixed by the elements of each column of the loading matrix $\boldsymbol{\Phi}_{r \times q}$ to approximate the corresponding column of $\boldsymbol{Y}$ [10, 8].

Assuming that $\boldsymbol{\Phi}$ is the projection of $\boldsymbol{Y}$ onto $\boldsymbol{W}$, i.e. $\boldsymbol{\Phi} = \boldsymbol{W}^T\boldsymbol{Y}$, the non-negativity constraint on the elements of $\boldsymbol{W}$ and $\boldsymbol{\Phi}$ makes our non-negative component analysis an optimization problem of minimizing the cost function $F(\boldsymbol{W}) = \|\boldsymbol{Y} - \boldsymbol{W}\boldsymbol{W}^T\boldsymbol{Y}\|^2$ with respect to $\boldsymbol{W}$, where $\|.\|$ represents the matrix norm. Considering the Frobenius norm, the optimization problem can be denoted by [10]

$$\min_{\boldsymbol{W} \geq 0} F_{fro}(\boldsymbol{W}) = \min_{\boldsymbol{W} \geq 0} trace\left\{ \left(\boldsymbol{Y} - \boldsymbol{W}\boldsymbol{W}^T\boldsymbol{Y}\right)\left(\boldsymbol{Y} - \boldsymbol{W}\boldsymbol{W}^T\boldsymbol{Y}\right)^T \right\}. \tag{2}$$

The above cost function can be minimized by a gradient descent approach, i.e. updating $W_{i,j} = W_{i,j} - \eta_{i,j}\frac{\partial F}{\partial W_{i,j}}$ with a positive step-size $\eta_{i,j}$ where

$$\frac{\partial F}{\partial W_{i,j}} = -4\left(\boldsymbol{Y}\boldsymbol{Y}^T\boldsymbol{W}\right)_{i,j} + 2\left(\boldsymbol{W}\boldsymbol{W}^T\boldsymbol{Y}\boldsymbol{Y}^T\boldsymbol{W}\right)_{i,j} + 2\left(\boldsymbol{Y}\boldsymbol{Y}^T\boldsymbol{W}\boldsymbol{W}^T\boldsymbol{W}\right)_{i,j}. \tag{3}$$

The gradient descent updating, as stated above, does not guarantee to keep the elements of $\boldsymbol{W}$ non-negative. However, due to the positivity of the elements of $\boldsymbol{Y}$ and by applying positive initialization to $\boldsymbol{W}$, our non-negativity constraint is guaranteed by having the step-size as follows

$$\eta_{i,j} = \frac{\frac{1}{2}W_{i,j}}{(\boldsymbol{W}\boldsymbol{W}^T\boldsymbol{Y}\boldsymbol{Y}^T\boldsymbol{W})_{i,j} + (\boldsymbol{Y}\boldsymbol{Y}^T\boldsymbol{W}\boldsymbol{W}^T\boldsymbol{W})_{i,j}}, \tag{4}$$

which results in the following multiplicative updating procedure [10]

$$W_{i,j} = W_{i,j}\frac{2\left(\boldsymbol{Y}\boldsymbol{Y}^T\boldsymbol{W}\right)_{i,j}}{(\boldsymbol{W}\boldsymbol{W}^T\boldsymbol{Y}\boldsymbol{Y}^T\boldsymbol{W})_{i,j} + (\boldsymbol{Y}\boldsymbol{Y}^T\boldsymbol{W}\boldsymbol{W}^T\boldsymbol{W})_{i,j}}. \tag{5}$$

For stability of the convergence, at each iteration, $\boldsymbol{W}$ is normalized by $\boldsymbol{W} = \frac{\boldsymbol{W}}{\|\boldsymbol{W}\|_2}$. Starting by initial random positive elements on $\boldsymbol{W}$, the iterative procedure will converge to the desired $\boldsymbol{W} \geq 0$ whose columns are PNCA. Each obtained component $\boldsymbol{w}_j$ (the jth column of $\boldsymbol{W}$) is then normalized by its norm $\|\boldsymbol{w}_j\|$,

and correspondingly the coefficients, i.e. elements of the matrix $\boldsymbol{\Phi} = [\varphi_{ji}]$, are adjusted as $\varphi_{ji} = \frac{\varphi_{ji}}{\|\boldsymbol{w}_j\|} : 1 < i < q$.

The resulting non-negative coefficients are an indicator of the weight of the corresponding component in reconstructing the original connectivity matrices. Therefore, we rank each component $\boldsymbol{w}_j$ based on the average of corresponding coefficients, i.e. $\frac{1}{q}\sum_{i=1}^{q}\varphi_{ji}$.

2.2   Group PNCA Model for SL Networks

As stated by (1), the q connectivity observations, i.e. $\boldsymbol{y}_i : \; 1 \leq i \leq q$, in the matrix $\boldsymbol{Y}$ are approximated by

$$[\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_q] \approx [\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_r] \begin{bmatrix} \varphi_{11} & \varphi_{12} & \cdots & \varphi_{1q} \\ \vdots & \vdots & & \vdots \\ \varphi_{r1} & \varphi_{r2} & \cdots & \varphi_{rq} \end{bmatrix}. \tag{6}$$

Each observation vector per subject i is thus, approximately reconstructed by

$$\boldsymbol{y}_i \approx \sum_{j=1}^{r}\varphi_{ji}\boldsymbol{w}_j = \sum_{j=1}^{r}\left(\boldsymbol{w}_j^T\boldsymbol{y}_i\right)\boldsymbol{w}_j \; ; \; 1 \leq i \leq q. \tag{7}$$

Let us suppose, with no loss of generality, that the first q_1 elements are from the first group (e.g. population of ASD) and the remaining from the second group (e.g. TD group). Thereby, the role of each component $\boldsymbol{w}_j$ in reconstructing the corresponding connectivity vector in the first group, i.e. $\boldsymbol{y}_i : \; 1 \leq i \leq q_1$, is characterized by the corresponding coefficients φ_{ji}; and so forth for the second group. Therefore, the statistical significance between the set of $\{\varphi_{ji} : \; 1 \leq i \leq q_1\}$ and $\{\varphi_{ji} : \; q_1 + 1 \leq i \leq q\}$ describes the importance of the corresponding basis connectivity component $\boldsymbol{w}_j$ in differentiating the two groups.

2.3   Synchronization Likelihood (SL) Connectivity Networks

Time-frequency synchronization likelihood assumes that two signals are highly synchronized if a pattern of one signal repeats itself at certain time instants within a time period while the other signal repeats itself at those same instants [9]. Such signal patterns at a time instant t_i of channel k can be represented by an embedding vector $\boldsymbol{x}_{k,t_i} = [x_k(t_i), x_k(t_{i+l}), \ldots, x_k(t_{i+(m-1)l})]$ where l is the lag and m is the length of the embedding vector. l and m are typically set to $l = \frac{f_s}{3h_f}$ and $m = \frac{3h_f}{l_f} + 1$ where f_s is the sampling frequency, and h_f and l_f are the high and low frequency contents of the signal, respectively [9]. At each time instant t_i, the Euclidean distance is measured between the reference embedding vector $\boldsymbol{x}_{k,t_i}$ and the set of all other embedding vectors at times t_j, i.e. $\boldsymbol{x}_{k,t_j}$, where t_j lies in the range $t_i - \frac{t_{w_2}}{2} < t_j < t_i - \frac{t_{w_1}}{2}$ or $t_i + \frac{t_{w_1}}{2} < t_j < t_i + \frac{t_{w_2}}{2}$ (in this work t_{w_2} was set to 10 sec and $t_{w_1} = \frac{2l(m-1)}{f_s}$). Then, n_{ref} nearest embedding vectors $\boldsymbol{x}_{k,t_j}$ are retained [9].

This procedure is conducted for each channel k and each time instants t_i. Then, the SL between channel k_1 and channel k_2 at time instant t_i is the number of simultaneous recurrences in the two channels divided by the total number of recurrences, i.e. $SL_{t_i} = n_{k_1 k_2}/n_{ref}$. The synchronization likelihood at all the time instants t_i are then averaged yielding the final SL between the two channels.

3   Results

3.1   Simulation Experiments

In order to demonstrate the effectiveness of the proposed method of PNCA in identifying the true underlying connectivity components, we apply our method to simulated connectivity matrices. We compare this with ICA which is the most widely-adopted method for similar purposes. The simulation is performed by using random non-negative numbers as the elements of three 10×10 simulated symmetric SL matrices plus a small background random (non-negative) noise. Ten linear mixtures of the simulated connectivity matrices are composed by a random mixing matrix with non-negative elements. The upper triangular part (excluding the diagonal) of the 10 connectivity matrices are vectorized to form the 10 columns of $\boldsymbol{Y}_{45 \times 10}$. We apply the fast ICA algorithm [11] as well as the proposed technique in Sect. 2.1 to solve for $r = 3$ normalized components as columns of $\boldsymbol{W}_{45 \times 3}$. For visualization, the SL matrices and the elements of the resulting components from fast ICA and PNCA algorithms are displayed by grayscale images. The results are shown in Fig. 1. It can be seen that while our method produces the right components (as compared to the original components in Fig. 1(a)), ICA is unable to resolve the components correctly. This is due to the fact that ICA forces the components to be statistically independent (i.e. with covariance of identity) whereas PNCA forces the components to be non-negative and orthogonal which leads to localized components due to sparsity. Please note that the resulting components are not ranked and ordered in Fig. 1.

3.2   PNCA on MEG SL Connectivity Networks

Dataset and preprocessing. Our dataset consisted of 48 children subjects (26 ASD's and 22 TD's) aged 6-15 years (mean=10.1, SD=2.3 in ASD, and mean=11.0, SD=2.5 in TD). A standard t-test showed that the populations' age difference was not statistically significant ($p - value > 0.19$). All subjects reported are free from medication use. Resting-state data were collected by a 275-channel MEG system, 274 channels being effective at the time of recording, with three head-position indicator coils to monitor head motion and ensure immobility. After a band-pass filter (0.03-150 Hz), MEG signals were digitized at 1200 Hz and downsampled offline to 500 Hz. Eye-blink artifacts were removed by learning the pattern of eye blinks and deleting that component from the data.

The recordings are then used to calculate the 274×274 SL matrices for delta (0.5–4 Hz) frequency band using a fourth order Butterworth bandpass filter. This

(a) Original components   **(b)** 10 mixtures of original components

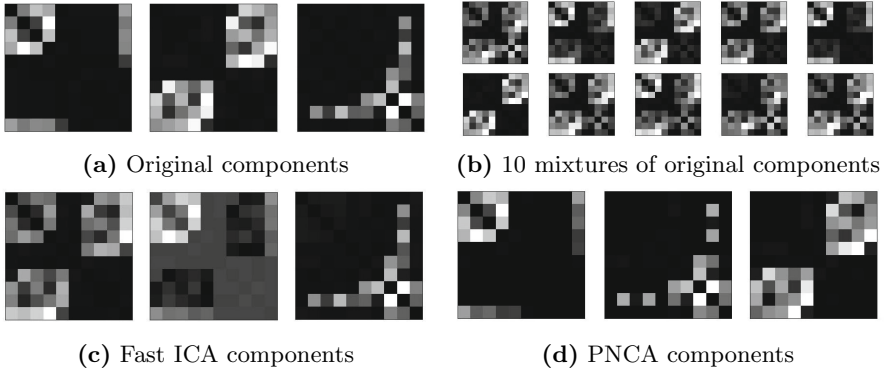**(c)** Fast ICA components   **(d)** PNCA components

Fig. 1. The results of fast ICA and PNCA on the simulated SL matrices. (a) The simulated SL matrices as the original components. (b) 10 linear mixtures of the components by random weightings. (c) The solution of fast ICA. (d) The solution of PNCA.

could be computed in any frequency band, but ASD studies, e.g. [1], have shown delta band anomalies and hence we concentrate on the delta band. Therefore, a total of 26 and 22 SL matrices are obtained for 26 ASD and 22 TD subjects, respectively. The upper triangular (excluding the diagonal) elements of each SL matrix are concatenated to make a vector of $\frac{274 \times 273}{2} = 37401$ elements which form one column of $\boldsymbol{Y}_{37401 \times 48}$.

Component Analysis. In order to determine the number of components used in the PNCA decomposition, we separately apply PCA to the aggregated connectivity vectors of populations of ASD, TD, and pooled ASD and TD (needed for a joint statistical analysis). The PCA findings indicated that a decomposition consisting of five ($r = 5$) components would account for 93% to 95% of the total variance. The PNCA is applied to the three cases of ASD ($\boldsymbol{Y}_{37401 \times 26}$), TD ($\boldsymbol{Y}_{37401 \times 22}$), and pooled ASD and TD ($\boldsymbol{Y}_{37401 \times 48}$), and the model of (1) is solved for five components ($\boldsymbol{W}_{37401 \times 5}$) for each of the three population cases. The resulting 37401-length connectivity components at each column of $\boldsymbol{W}$ are then used to form the corresponding 274×274 symmetric connectivity matrices. Figure 2 shows the resulting five connectivity components on the MEG sensor map. These components are ranked here (left to right) based on the descending average of their corresponding coefficients in each of the three population cases. The averages of the coefficients are given in Table 1.

A statistical group analysis, as described in Sect. 2.2, was performed over the resulting PNCA coefficients, i.e. φ_{ji}. The two-sample t-test is applied to the coefficients of each normalized connectivity component from the pooled AST–TD and the $p-$values and $t-$values are given in Table 1.

We see that the average of the first component weights (φ_{1i}) are statistically smaller in constructing the SL connectivity matrices of the ASD group compared to the TD group. Interestingly, this is the most dominate component, ranked by the average of the corresponding coefficients, in the pooled

(a) Connectivity components of ASD

(b) Connectivity components of TD

(c) Connectivity components of pooled ASD and TD
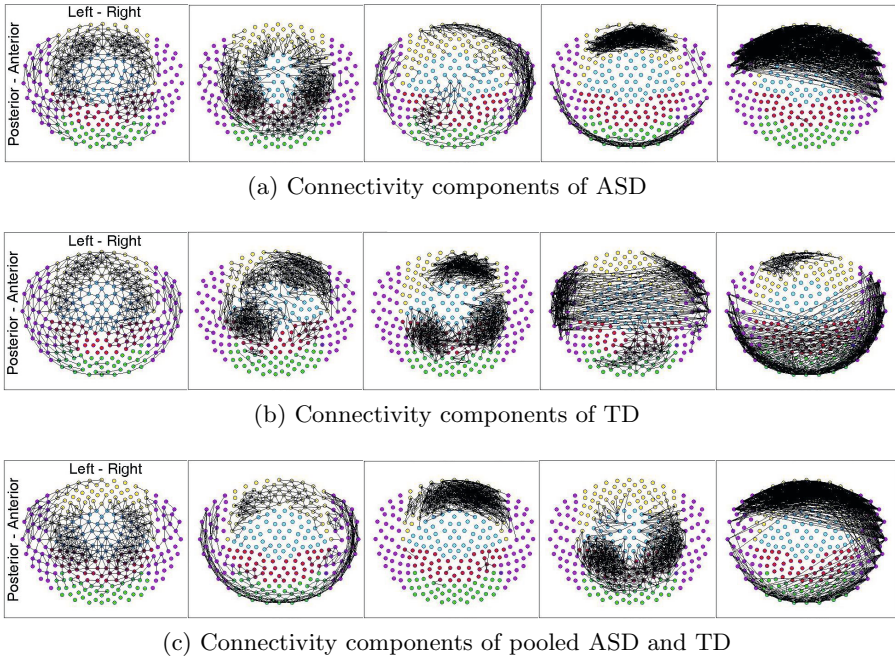
Fig. 2. PNCA connectivity components plotted on the MEG sensor map and sorted based on their average coefficients

Table 1. PNCA component ranking and statistical group analysis

Component No.	ASD Average Coeff's	TD Average Coeff's	Pooled ASD-TD Average Coeff's	ASD-TD group $p-$value	ASD-TD group $t-$value
1	10.2	11.4	9.7	0.02	-2.3
2	6.2	6.4	7.0	0.32	-1.0
3	6.0	4.8	5.7	0.75	+0.3
4	4.6	4.6	5.0	0.28	-1.1
5	2.0	2.8	2.8	0.23	+1.2

population, and is quite similar to the most dominant components determined from the separate ASD and TD populations. This component can be interpreted as the default connectivity network observed in the resting brain. It is also notable that the presence of the fourth and fifth components of ASD indicates strong short range frontal connectivity diminished in TD, while the fourth component in TD population, with clear long range connectivity, is not found in the ASD set of components. Together these support intact default connectivity in ASD with evidence for diminished long range and enhanced short-range connectivity in ASD; a finding consistent with other connectivity analysis investigations in autism [1].

4   Conclusion

We have presented a non-negative component analysis technique for learning localized and part-based sparse components of non-negative connectivity matrices. The algorithm is based on non-negative projections which produces non-negative bases and coefficients by a gradient descent approach minimizing a Frobenius norm of the reconstruction matrix error. We applied it to the simulated connectivity matrices which showed more accurate component findings compared to the well-known ICA technique. The proposed method is then applied to the novel problem of investigating MEG derived SL connectivity matrices, within a group study of ASD. The projections of the connectivity elements onto the components revealed statistically significant differences between how the ASD and TD functional connectivity matrices are composed of their fundamental connectivity components. The presented technique represents a framework, in principle capable of handling other types of functional or structural connectivity networks from any modality for statistical group analysis or feature selection.

References

1. Barttfeld, P., et al.: A big-world network in asd: Dynamical connectivity analysis reflects a deficit in long-range connections and an excess of short-range connections. Neuropsychologia 49, 254–263 (2011)
2. Skudlarski, P., et al.: Brain connectivity is not only lower but different in schizophrenia: a combined anatomical and functional approach. Biol. Psychiatry 68(1), 61–69 (2010)
3. Tsiaras, V., et al.: Extracting biomarkers of autism from meg resting-state functional connectivity networks. Comput. Biol. Med. 41(12), 1166–1177 (2011)
4. Vissers, M., Cohen, M., Geurts, H.: Brain connectivity and high functioning autism: a promising path of research that needs refined models, methodological convergence, and stronger behavioral links. Neurosci. Biobehav. Rev. 36(1), 604–625 (2012)
5. Calhoun, V., Kiehl, K., Pearlson, G.: Modulation of temporally coherent brain networks estimated using ica at rest and during cognitive tasks. Hum. Brain Mapp. 29(7), 828–838 (2008)
6. Calhoun, V., Adali, T., Pearlson, G., Pekar, J.: A method for making group inferences from functional mri data using independent component analysis. Hum. Brain Mapp. 14(3), 140–151 (2001)
7. Yang, Z., Yuan, Z., Laaksonen, J.: Projective non-negative matrix factorization with applications to facial image processing. Int. J. of Pattern Recog. and Artif. Intell. 21(8), 1353–1362 (2007)
8. Batmanghelich, N., Taskar, B., Davatzikos, C.: Generative-discriminative basis learning for medical imaging. IEEE Trans. Med. Imaging 31(1), 51–69 (2011)
9. Montez, T., Linkenkaer-Hansen, K., van Dijk, B., Stam, C.: Synchronization likelihood with explicit time-frequency priors. Neuroimage 33(4), 1117–1125 (2006)
10. Yang, Z., Oja, E.: Linear and nonlinear projective nonnegative matrix factorization. IEEE Trans. Neural Netw. 21(5), 1734–1749 (2010)
11. Hyvarinen, A., Oja, E.: Independent component analysis: algorithms and applications. Neural Netw. 13(4-5), 411–430 (2000)

Tree-Guided Sparse Coding for Brain Disease Classification[*]

Manhua Liu[1,2], Daoqiang Zhang[1,3], Pew-Thian Yap[1], and Dinggang Shen[1]

[1] IDEA Lab, Department of Radiology and BRIC,
University of North Carolina at Chapel Hill, USA
[2] Department of Instrument Science and Technology,
Shanghai Jiao Tong University, China
[3] Department of Computer Science and Engineering,
Nanjing University of Aeronautics and Astronautics, China
`dgshen@med.unc.edu`

Abstract. Neuroimage analysis based on machine learning technologies has been widely employed to assist the diagnosis of brain diseases such as Alzheimer's disease and its prodromal stage - mild cognitive impairment. One of the major problems in brain image analysis involves learning the most relevant features from a huge set of raw imaging features, which are far more numerous than the training samples. This makes the tasks of both disease classification and interpretation extremely challenging. Sparse coding via L1 regularization, such as Lasso, can provide an effective way to select the most relevant features for alleviating the curse of dimensionality and achieving more accurate classification. However, the selected features may distribute randomly throughout the whole brain, although in reality disease-induced abnormal changes often happen in a few contiguous regions. To address this issue, we investigate a tree-guided sparse coding method to identify grouped imaging features in the brain regions for guiding disease classification and interpretation. Spatial relationships of the image structures are imposed during sparse coding with a tree-guided regularization. Our experimental results on the ADNI dataset show that the tree-guided sparse coding method not only achieves better classification accuracy, but also allows for more meaningful diagnosis of brain diseases compared with the conventional L1-regularized Lasso.

1 Introduction

Neuroimaging data, such as magnetic resonance image (MRI) and fluorodeoxyglucose positron emission tomography (FDG-PET), provides a powerful in vivo tool for aiding diagnosis and monitoring of brain diseases, such as Alzheimer's disease (AD) and mild cognitive impairment (MCI) [1, 2]. Recently, many machine learning and pattern

---

recognition technologies, e.g., support vector machines (SVM), have been investigated for analysis of brain images to assist the diagnosis of brain diseases [1-6]. However, the original neuroimaging data of the whole brain is often of huge dimensionality, and their direct use for control-patient classification is not only computationally expensive, but also could lead to low performance since not all features are relevant to disease pathology. Thus, feature extraction and selection are necessary and important for identifying the most relevant and discriminative features for guiding classification.

Morphological analysis of brain images has been widely used to investigate the pathological changes related to the brain diseases. One popular method is to group voxels into multiple anatomical regions, i.e., regions of interest (ROIs), through the warping of a pre-labeled atlas, and then extract regional features such as anatomical volumes for guiding the classification [1, 7, 8]. However, this approach to anatomical parcellation may not adapt well to the diseased-related pathology since the abnormal region may be part of ROI or span over multiple ROIs. To address this issue, Fan *et al.* [9] proposed to adaptively partition the brain image into a number of most discriminative brain regions according to the similarity computed based on correlation of image features with respect to the class labels. Then, regional features were extracted for brain disease classification. In addition to significantly reduce the feature dimensionality, this method is also robust to noise and registration error. However, the extracted regional features are generally very coarse and not sensitive to small local changes, thus affecting classification performance. Although this limitation could be potentially solved by voxel-wise analysis method [10], i.e., using voxel-wise features for classification, the number of voxel-wise features from the whole brain is often very large (i.e., in millions), while the number of training samples is very small (i.e., in hundreds) in the neuroimaging study. This could also cause a significant drop in performance for high-dimensional classification methods, such as support vector machines (SVM) [11]. Therefore, it is important to significantly reduce the number of voxel-wise features before performing classification.

So far, many feature reduction and selection techniques have been proposed to select a small number of discriminative features for brain classification. Principal Component Analysis (PCA) is one of the popular methods to reduce the feature space to the most discriminant components [12]. It performs a linear transformation of the data to a lower dimensional feature space for maximization of data variance, and thus cannot always detect features from those localized abnormal brain regions. Another popular method is to select the most discriminative features and eliminate the redundant features in terms of the correlations of the individual features to the group difference such as *t*-test [3]. However, this selection method does not consider the relationships of imaging features, thus limiting its ability to detect the complex population difference.

Recently, L1-regularized sparse coding methods, e.g., Lasso, was proposed and used to sparsely identify a small subset of input features to best represent the outputs [13], and promising results were obtained. However, the selected features by L1-regularization may distribute randomly throughout the whole brain, although in reality

the disease-induced abnormal changes often happen in a few number of contiguous brain regions, instead of isolated voxels. This makes the interpretation of classification results very difficult. Actually, spatially adjacent voxels of a brain image are usually correlated, thus the coefficients assigned to them during the L1-regularization should have similar magnitudes in order to reflect their underlying correlations. Recently, a group sparse coding (Lasso) method with the hierarchical tree-guided regularization was proposed as an extension of Lasso to consider the underlying structural information among the inputs or outputs [14, 15]. In this paper, we propose to apply this tree-guided group Lasso method to identify the relevant biomarkers with the structured sparsity from MR images for brain disease classification. The hierarchical relationships of the imaging features in the whole brain are imposed in the regularization of sparse coding by a tree structure. Our experimental results on ADNI database demonstrate that, in addition to better classify the neuroimaging data of AD and MCI, the proposed classification algorithm can also identify the structured relevant biomarkers to facilitate the interpretation of classification results.

2   Method

Assume we have M training brain images, with each represented by a N-dimensional feature vector and a respective class label. The classification problem involves selection of the most relevant features and also decoding the disease states of the images, i.e., the class labels. It is observed that there are only a few brain regions affected by the disease. Thus, sparsity can be incorporated into the learning model for feature selection and disease classification.

2.1   L1-Regularized Sparse Coding (Lasso)

Let X denote a $N{\times}M$ feature matrix with the m-th column corresponding to the m-th image's feature vector $x_m = (x_m^1, \dots, x_m^n, \dots, x_m^N)^T \in R^N$ and y be a class label vector of M images with y_m denoting the class label of the m-th image. A linear model can be assumed to decode the class outputs from a set of features as follows:

$$y = X\alpha + \varepsilon \tag{1}$$

where $\alpha = (\alpha_1, \dots, \alpha_n, \dots, \alpha_N)^T$ is a vector of coefficients assigned to the respective features, and ε is an independent error term. The least square optimization is one of the popular methods to solve the above problem. When N is large and the number of features relevant to the class labels is small, sparsity can be imposed on the coefficients of the least square optimization via L1-norm regularization for feature selection [13, 16]. The L1-regularized least square problem, i.e., Lasso, can be formulated as:

$$\alpha = arg\min_{\alpha} \|y - X\alpha\|^2 + \lambda\|\alpha\|_1 \tag{2}$$

where λ is a regularization parameter that controls the amount of sparsity in the solution. The non-zero elements of α indicate that the corresponding input features are relevant to the class labels.

The L1-regularized sparse coding provides an effective way to select a small subset of features by taking into account the correlations of individual features to the class labels. However, the structural relationships among the features, which are an important source of information, are ignored in this method. In some situations, the associated features should be jointly selected to identify the complex population difference. For example, the disease-induced abnormal changes often happen in the contiguous regions of brain image, instead of isolated voxels.

2.2   Tree-Guided Sparse Coding

To reduce the feature dimensionality while taking into account the structural relationships among the features, group Lasso was proposed as an extension of Lasso to use the groups of features instead of individual features as the units of feature selection [14]. In the regularization of sparse coding, group Lasso applies L1-norm penalty over the feature groups and L2-norm penalty for the features within each group. It assumes that the groupings of features are available as prior knowledge. However, in practice, a prior knowledge about the structures and relationships among the brain imaging features is not always available. In many applications, the features can be naturally represented using a tree structure to reflect their hierarchical spatial relationships. A tree-guided group lasso was proposed for multi-task learning where multiple related tasks follow a tree structure [14].

The brain image shows spatial correlations between the neighboring voxels, forming groups of different sizes and shapes. In this work, we propose to apply a tree structure to represent the hierarchical spatial relationships of brain image structure, with leaf nodes as the imaging features and internal nodes as the groups of features. A regularization predefined by the tree structure can be imposed on the sparse coding optimization problem to encourage a joint selection of structured relevant features. Fig. 1 shows a hierarchical tree structure imposed on a sample brain image. Assume that an index hierarchical tree T of d depth levels with $T_i = \{G_1^i, \dots, G_j^i, \dots, G_{n_i}^i\}$ containing n_i nodes in the ith level, $0 \leq i \leq d$. The different depth levels indicate the variant scales of feature groups. The index sets of the nodes in the same level have no overlapping while the index sets of a child node is a subset of its parent node. The tree-guided group Lasso (sparse coding) method can be formulated as:

$$\alpha = arg \min_{\alpha} \|y - X\alpha\|^2 + \lambda \sum_{i=0}^{d} \sum_{j=1}^{n_i} w_j^i \left\| \alpha_{G_j^i} \right\|_2 \tag{3}$$

where $\alpha_{G_j^i}$ is the set of coefficients assigned to the features within node G_j^i, w_j^i is a predefined weight for node G_j^i and is usually set to be proportional to the square root of the group size, and the number of depth levels d is set to 3 as in Fig.1. The features with non-zero coefficients are finally selected for further classification.
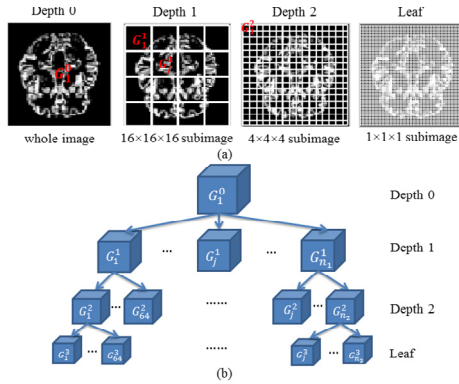
Fig. 1. Illustration of the tree structure using 2D slice as an example: (a) the subimages in different levels of tree and (b) the hierarchical tree nodes and leaves

2.3   Classification

Based on the selected imaging features by the tree-guided sparse coding method, a classifier model will be trained to make the final classification. There are various classifier models investigated for classification of brain diseases. Among them, SVM is one of the widely used classifiers because of its high classification performance [1, 7, 9, 12]. SVM constructs a maximal margin classifier in a high-dimensional feature space by mapping the original features using a kernel-induced mapping function. We choose the SVM model with a linear kernel and implement it using MATLAB SVM toolbox and the default parameters to train a classifier with the selected features for classification.

3   Experiments

We evaluate the proposed classification algorithm with the T1-weighted baseline MR brain images of 643 subjects, which include 196 AD patients, 220 MCI subjects, and 227 normal controls (NC), randomly selected from Alzheimer's Disease Neuroimaging Initiative (ADNI) database. Table 1 provides a summary of the demographic characteristics of the studied subjects (denoted as mean ± standard deviation). Before performing classification, the image preprocessing was performed as follows. All MR brain images were first skull-stripped and cerebellum-removed after a correction of intensity inhomogeneity [17]. Then, each image was segmented into three brain tissues, i.e., gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF), which were spatially normalized onto a standard space by a mass-preserving deformable registration algorithm [18]. The spatially normalized tissues are called as tissue densities in this paper. To reduce the effects of noise, registration inaccuracy, and inter-individual anatomical variations, tissue density maps were further smoothed using a Gaussian filter and then down-sampled by a factor for the purpose of saving the computational time.

Table 1. Demographic characteristics of the studied subjects from ADNI database

Diagnosis	Number	Age	Gender (M/F)	MMSE
AD	196	75.6±7.7	103/93	23.2±2.0
MCI	220	75.0±7.4	150/70	26.8±1.8
NC	227	76.1±5.0	117/110	29.0±1.0

In the experiments, we only use the GM density map as the imaging features because it is more relevant to AD and MCI. The tree-guided group Lasso is implemented using SLEP (http://www.public.asu.edu/~jye02/Software/SLEP). The proposed algorithm is performed to classify AD vs NC and MCI vs NC. To evaluate the classification performance, 10-folds cross-validations were performed to compute the classification accuracy, i.e., the proportion of correctly classified subjects among the test dataset. In addition, we also test the L1-regularized sparse coding (Lasso) on the same dataset for comparison. Specifically, the classification accuracies are compared with respect to different number of the selected features by sparse coding. In the experiments, we change the regularization parameter $\lambda \in [0\ 1]$ to adjust the sparsity and obtain different number of selected features, with the increasing of λ leading to a smaller number of features. The classification results for AD vs NC and MCI vs NC are shown in Fig. 2. From these results, we can see that the tree-guided group Lasso can achieve better classification accuracy than the L1-regularized Lasso when the number of selected features is small ($<1.4 \times 10^4$, i.e., less than half of available features). The classification results have no large difference for both methods when the number of selected features is large. This indicates that further increasing the number of selected features will reduce the effect of structure constraint of sparse coding in classification. Also, the low computation time (i.e., needing less than 3 seconds for feature selection) makes it feasible for grouping of effective voxel-wise features.
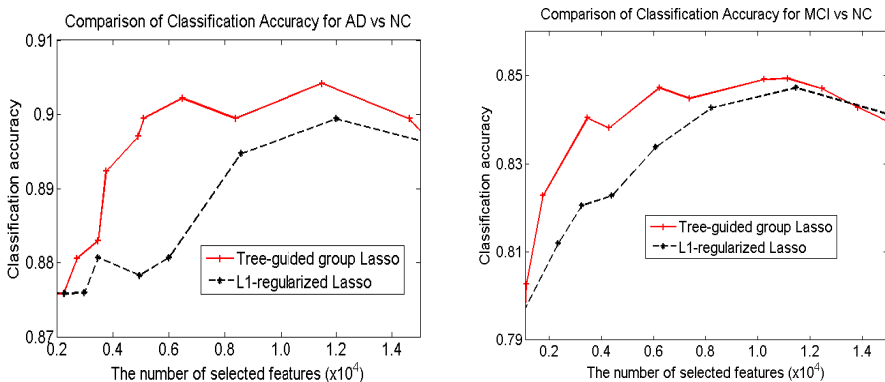
Fig. 2. Comparison of classification accuracies by the tree-guided group Lasso and L1-regularized Lasso, with respect to different number of selected features

(a)

(b)
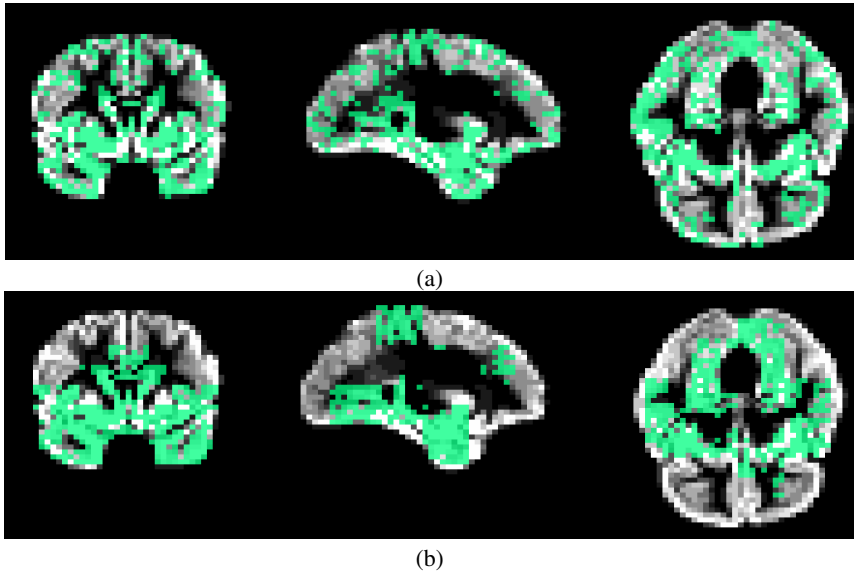
Fig. 3. The imaging features identified from the GM density map by (a) L1-regularized Lasso and (b) tree-guided group Lasso, for the case of classifying AD vs NC

(a)

(b)
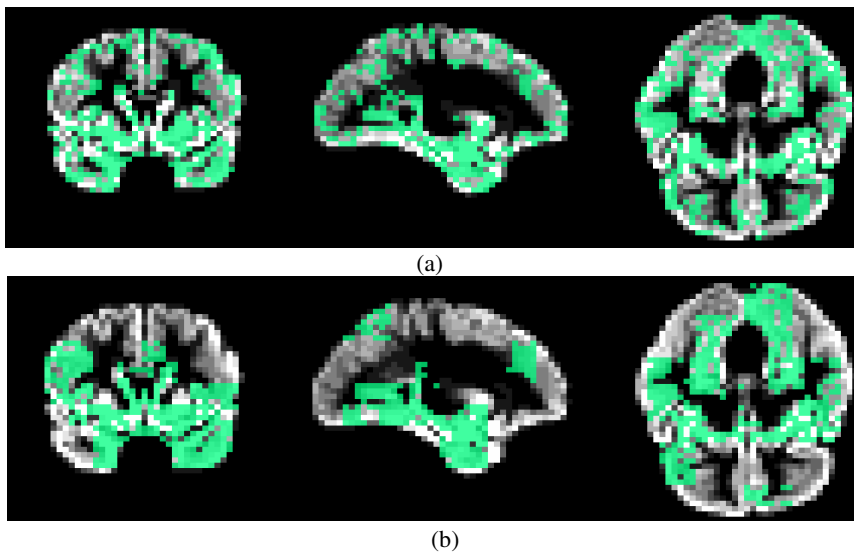
Fig. 4. The imaging features identified from the GM density map by (a) L1-regularized Lasso and (b) tree-guided group Lasso, for the case of classifying MCI vs NC

For interpretations, we show the selected image features by both the L1-regularized and tree-guided Lasso methods, with their own best regularization parameters, in Fig. 3 and 4 for AD vs NC and MCI vs NC classifications, respectively. We can see that the spatial overlaps between the L1-regularized and tree-guided Lasso methods are usually at the most relevant regions such as hippocampus, entorhinal cortex, and parahippocampal gyrus. But the features selected by L1-regularized Lasso are irregularly distributed throughout the whole brain, while the features selected by tree-guided Lasso are usually grouped at the relevant regions which are able to facilitate the interpretation of the obtained results. We evaluated that the resulting regions identified by tree-guided Lasso include hippocampus, entorhinal cortex, parahippocampal gyrus, and amygdala, which are consistent with those reported in the literature for AD and MCI studies [4, 5, 7]. These results verify the effectiveness of the tree-guided sparse coding method in incorporating the spatial structure and relationships of imaging features for guiding the disease classification and also identification of grouped relevant features.

4   Conclusion

In this paper, a sparse coding method with a tree-guided regularization is investigated to sparsely identify the grouped relevant biomarkers for brain disease classification. The tree-guided regularization is used to capture the hierarchical spatial relationships among the imaging features. Thus, the tree-guided sparse coding can provide an effective way to identify the meaningful biomarkers for brain disease classification and interpretation. Experimental results on ADNI dataset show that the proposed method not only identifies the grouped relevant biomarkers but also achieves better classification performance than the conventional L1-regularized Lasso method. Although we test this method only on classification of MR images for AD and MCI diagnosis, the similar idea can be extended and applied to other neuroimaging modalities for diagnosis of AD or other brain diseases.

References

1. Magnin, B., Mesrob, L., Kinkingnehun, S., Pelegrini-Issac, M., Colliot, O., Sarazin, M., Dubois, B., Lehericy, S., Benali, H.: Support vector machine-based classification of Alzheimer's disease from whole-brain anatomical MRI. Neuroradiology 51, 73–83 (2009)
2. Wolz, R., Julkunen, V., Koikkalainen, J., Niskanen, E., Zhang, D.P., Rueckert, D., Soininen, H., Lötjönen, J.: Multi-Method Analysis of MRI Images in Early Diagnostics of Alzheimer's Disease. PloS One 6, e25446 (2011)
3. Davatzikos, C., Fan, Y., Wu, X., Shen, D., Resnick, S.M.: Detection of Prodromal Alzheimer's Disease via Pattern Classification of MRI. Neurobiol. Aging (2006, epub)
4. Hinrichs, C., Singh, V., Mukherjee, L., Xu, G., Chung, M.K., Johnson, S.C.: Spatially augmented LPboosting for AD classification with evaluations on the ADNI dataset. Neuroimage 48, 138–149 (2009)

5. 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 56, 766–781 (2011)

6. Vemuri, P., Gunter, J.L., Senjem, M.L., Whitwell, J.L., Kantarci, K., Knopman, D.S., Boeve, B.F., Petersen, R.C., Jack Jr., C.R.: Alzheimer's disease diagnosis in individual subjects using structural MR images: Validation studies. Neuroimage 39, 1186–1197 (2008)

7. Zhang, D., Wang, Y., Zhou, L., Yuan, H., Shen, D.: Multimodal classification of Alzheimer's disease and mild cognitive impairment. Neuroimage 55, 856–867 (2011)

8. Lao, Z., Shen, D., Xue, Z., Karacali, B., Resnick, S.M., Davatzikos, C.: Morphological classification of brains via high-dimensional shape transformations and machine learning methods. Neuroimage 21, 46–57 (2004)

9. 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, 93–105 (2007)

10. Ishii, K., Kawachi, T., Sasaki, H., Kono, A.K., Fukuda, T., Kojima, Y., Mori, E.: Voxel-based morphometric comparison between early- and late-onset mild Alzheimer's disease and assessment of diagnostic performance of z score images. American Journal of Neuroradiology 26, 333–340 (2005)

11. Chapelle, O., Vapnik, V., Bousquet, O., Mukherjee, S.: Choosing multiple parameters for support vector machines. Mach. Learn. 46, 131–159 (2002)

12. Davatzikos, C., Resnick, S.M., Wu, X., Parmpi, P., Clark, C.M.: Individual patient diagnosis of AD and FTD via high-dimensional pattern classification of MRI. Neuroimage 41, 1220–1227 (2008)

13. Ghosh, D., Chinnaiyan, A.M.: Classification and selection of biomarkers in genomic data using LASSO. J. Biomed. Biotechnol., 147–154 (2005)

14. Kim, S., Xing, E.P.: Tree-guided group lasso for multi-task regression with structured sparsity. Arxiv preprint arXiv:0909.1373 (2009)

15. Liu, J., Ye, J.: Moreau-Yosida regularization for grouped tree structure learning. In: Advances in Neural Information Processing Systems (2010)

16. Tibshirani, R.: Regression shrinkage and selection via the Lasso. J. Roy. Stat. Soc. B Met. 58, 267–288 (1996)

17. 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)

18. Shen, D., Davatzikos, C.: Very high resolution morphometry using mass-preserving deformations and HAMMER elastic registration. Neuroimage 18, 28–41 (2003)

Improving Accuracy and Power with Transfer Learning Using a Meta-analytic Database

Yannick Schwartz[1,2], Gaël Varoquaux[1,2], Christophe Pallier[3,2], Philippe Pinel[3,2], Jean-Baptiste Poline[2], and Bertrand Thirion[1,2]

[1] Parietal Team, INRIA Saclay-Île-de-France, Saclay, France
`yannick.schwartz@inria.fr`
[2] CEA, DSV, I[2]BM, Neurospin bât 145, 91191 Gif-Sur-Yvette, France
[3] INSERM, CEA, Cognitive Neuroimaging Unit, Neurospin Center, France

Abstract. Typical cohorts in brain imaging studies are not large enough for systematic testing of all the information contained in the images. To build testable working hypotheses, investigators thus rely on analysis of previous work, sometimes formalized in a so-called *meta-analysis*. In brain imaging, this approach underlies the specification of regions of interest (ROIs) that are usually selected on the basis of the coordinates of previously detected effects. In this paper, we propose to use a database of images, rather than coordinates, and frame the problem as *transfer learning*: learning a discriminant model on a reference task to apply it to a different but related new task. To facilitate statistical analysis of small cohorts, we use a sparse discriminant model that selects predictive voxels on the reference task and thus provides a principled procedure to define ROIs. The benefits of our approach are twofold. First it uses the reference database for prediction, *i.e.* to provide potential biomarkers in a clinical setting. Second it increases statistical power on the new task. We demonstrate on a set of 18 pairs of functional MRI experimental conditions that our approach gives good prediction. In addition, on a specific transfer situation involving different scanners at different locations, we show that voxel selection based on transfer learning leads to higher detection power on small cohorts.

Keywords: Meta-analysis, fMRI, multiple comparison, machine learning.

1 Introduction

Multi-subject or multi-condition experiments are the workhorse of bio-medical imaging research, whether it be drug development or basic research. Imaging provides a wealth of information on the biomedical problem at hand. However the typical sample size is too small to fully exploit this information. For this reason, investigators often turn to previous studies in order to formulate hypotheses and restrict the search space, *i.e.* select a subset of anatomical or functional structures of interest to the current study. A typical case is that of early-stage clinical trials, for which the group size is very small, but that are most often based on previous results concerning the pathology under study. However, understanding the literature is increasingly difficult and requires a systematic approach, that

takes the form of a *meta-analysis*, pooling results from multiple experiments that address a set of related research hypotheses [1].

In particular, brain imaging studies heavily rely on such meta-analyses [2], as the brain is still an ill-understood and complex organ. In functional Magnetic Resonance Imaging (fMRI) studies, typical group sizes range from 10 to 20 subjects, which is not always enough to warrant the reliability of brain-wide analysis [3]. More importantly, the time that can be spent in the scanner by subjects is limited, and not all interesting experimental conditions will be acquired.For this reason, it is common practice to reduce the study to a set of regions of interest (ROIs) extracted from the literature. Investigators define these ROIs by extracting locations of peak activations from the literature [4], or from coordinate databases such as BrainMap [5]. While most of these meta-analyses are conducted on activation coordinates, the increase of data sharing opens the door to meta-analysis on full brain images which results in higher statistical power [6]. Previous statistical and modeling work on meta-analysis for fMRI has focused on better modeling of the reference database [2].

In this work, we are interested in the generalization power of meta-analyses on new data. We introduce a new meta-analysis method using a reference database of images to guide statistical analysis of a new dataset. In particular we rely on predictive models, useful to learn biomarkers, and use them to select relevant voxels in order to increase the statistical power of a new study.

2   Methods

Problem setting. We start from a reference database made of l experiments, each comprising n^l contrasts possibly acquired over multiple subjects. We denote the brain images by $\boldsymbol{X}^l \in \mathbb{R}^{n^l,p}$ with associated experimental condition $\boldsymbol{y}^l \in \mathbb{R}^{n^l}$. Given a new experiment, denoted *target*, $(\boldsymbol{X}^\star, \boldsymbol{y}^\star) \in (\mathbb{R}^{n^\star,p}, \mathbb{R}^{n^\star})$, we are interested in two problems: i) *(biomarkers)* can we predict $\boldsymbol{y}^\star$ from $\boldsymbol{X}^\star$? ii) *(inference)* can we test hypotheses on the links between $\boldsymbol{y}^\star$ and $\boldsymbol{X}^\star$, for instance in a linear model? These are ill-posed problems from the statistics standpoint, as $n^\star \ll p$. The root of the problem is the dimensionality of the data: medical images are composed of many voxels, typically $p \approx 50\,000$ with fMRI. This large number of descriptors limits statistical inference power due to multiple testing; a problem that appears in predictive approaches as the curse of dimensionality. Here, we use our reference database to better condition this statistical problem.

Transfer learning. The gist of our approach is to learn on some experiments of our database $(\boldsymbol{X}^l, \boldsymbol{y}^l)$ discriminative models that contain predictive information for the target experiment $(\boldsymbol{X}^\star, \boldsymbol{y}^\star)$. In machine learning, this problem is known as *transfer learning* [7]. The underlying assumption of transfer learning is the same as that for meta-analysis: the reference database should contain some common information with the target experiment. Here we use a simple form of transfer learning: we train a linear classifier on an experiment in the database that is similar from the neuroscientific point of view to the new data, and use it to predict the labels of the new data.

Selecting predictive features. We use a sparse linear classifier, specifically an ℓ_1-penalized logistic regression. The motivation behind this choice of classifier is that it produces a sparse set of weights that can be used to select relevant voxels. In particular, under certain conditions, the classifier can recover with high probability the complete set of k features in $\boldsymbol{X}$ that are predictive of $\boldsymbol{y}$ for a sample size of $n_{\min} = \mathcal{O}(k \log p)$ [8]. The logarithmic dependence in p is an appealing property in view of the dimensionality of medical imaging datasets.

In practical situations, it can be hard to control the errors on this feature selection, in particular as it depends on the choice of the amount of ℓ_1 penalty. For this reason, Meinshausen and Bühlmann [9] introduce randomized variants of sparse estimators, that can be seen as sampling the posterior probability of selection and keeping only features that are selected frequently. In particular, they establish non-asymptotic recovery results for the *randomized lasso*, which consists in applying the Lasso on random subsamples of the data and rescaling of the regressors. Here, we adapt this strategy to classification as the logistic regression is locally equivalent to a weighted least square and recovery results can carry from square-loss regression to logistic regression [8].

We want to use transfer learning to perform screening of the voxels, *i.e.* eliminate many voxels that are not related to our target experiment. For this purpose, we need a low probability of rejecting relevant variables. Each iteration of the sparse logistic regression in the randomized logistic can select reliably only $k_{\max} \approx n/\log p$ variables. In the worst case situation, we have k heavily-correlated variables and one of them is selected at random by the sparse logistic regression at each iteration. For each of these variables, the probability of selecting it less than s times during m iterations of the randomized logistic is given by the cumulative distribution function of a binomial with per trial success ratio $1/k$. If $s \leq m/k$, by Hoeffding's inequality, this probability goes to zero in $o(\exp m)$. In other words, if we impose a threshold $\tau = s/m$ on the selection frequency, we can recover a group of k correlated variables for $\tau \leq 1/k$.

Brain parcellations. Although randomization relaxes the conditions on recovery, a remaining necessary condition is that the regressors of interest, *i.e.* the values $\boldsymbol{x}_i$ across the subjects of the k predictive voxels, must be weakly correlated[1]. Because of the large amount of smoothness present in medical images, in particular in group-level fMRI contrasts, these conditions cannot be satisfied. Indeed, values taken by a voxel are very similar to values taken by its neighbors. In addition, the numbers of subjects used in fMRI are often below the sample size required for good recovery. For these reasons we resort to feature agglomeration: using hierarchical clustering to merge neighboring voxels carrying similar information into parcels [10]. This strategy brings the double benefit of reducing the problem size, and thus the required sample size, and mitigating local correlation, at the expense of spatial resolution.

---

[1] Specifically, the condition for recovery with randomized lasso it is a lower bound on the conditioning of the sparse eigenvalues of the design matrix [9, theorem 2] and for sparse logistic regression the corresponding condition is a lower bound on the eigenvalues of the regressors of interest's covariance matrix [8, theorem 4].

3   Experiments and Results

3.1   FRMI Datasets

We use 3 studies for this meta-analysis. The first study (*E1*) [11] is composed
of 322 subjects and was designed to assess the inter-subject variability in some
language, calculation, and sensorimotor tasks. The second study (*E2*) is similar
to the first one in terms of stimuli, but its data was acquired on 35 pairs of twin-
subjects. The last study (*E3*) [12] characterizes brain regions in charge of the
syntactic and the semantic processing for the language. It was performed with
40 subjects, 20 of which were stimulated by pseudowords (jabberwocky stimuli)
instead of actual meaningful sentences. All the studies were pre-processed and
analyzed with the standard fMRI analysis software SPM5. The data used for this
work are the statistical images resulting from the intra-subject analyses across
the 3 studies. *E1* has 34 contrasts images available, *E2* 56, and *E3* 28. The raw
images were not always acquired on the same scanner. *E1* has data from a 3T
SIEMENS Trio, and a 3T Brucker scanner; *E2* data were acquired on a 1.5T
GE Signa; and *E3* images come from the same 3T SIEMENS Trio.

3.2   Experimental Results for Prediction

Here we are interested in the *prediction* problem: using transfer learning to dis-
criminate a pair of contrasts with an estimator trained on two other contrasts.

We used 4 different approaches to learn the discriminative models. The first
approach relies on the activation likelihood estimate (ALE) method [13], as this
is a commonly published method for coordinate-based meta-analyses. We extract
the activation positions from the contrasts maps, and then apply a Gaussian ker-
nel. We use the preferred FWHM of 10mm [14]. The other approaches directly
use the contrast images. We name *raw contrasts* the method based on the con-
trasts voxels values; *contrast-specific parcels* the method that uses parcels from
the training set: and *meta-analystic parcels* the method that learns parcels from
the full database. We evaluate on our base of contrasts the ability to do transfer
learning, *i.e* to learn decision rules that carry over from one situation to another.
Since we must make the assumption that the reference contrasts hold common
information with the contrasts of interest, we do not try out all the possible com-
binations, but rather manually select pairs of contrasts from a single experiment
that form a meaningful classification task (*e.g.*, computation versus reading,
or Korean language versus French language). Out of all the possible combina-
tions, we select 35 pairs of classification task, and subsequently combine them
into 18 transfer pairs, on which it is reasonable to think that the transfer could
occur (*e.g.*, computation and reading in visual instructions, transfer on com-
putation and reading in auditory instructions). We first train a linear classifier
within 6-fold cross validation test on a first set of pairs, setting the penalization
amount by nested cross-validation, we call this step *inline learning*. We then
re-use the discriminative model on a different pair of contrasts to perform the
transfer learning. The 3 studies containing language related tasks, we can trans-
fer between pairs within an experiment, and across experiments. Among the 18

selected transfer pairs, we find that 9 can give rise to such a transfer. Since a transfer is directed, we perform it both ways, which yields once again 18 transfer pairs to test upon. The associated prediction scores from the different methods are reported in Table 1. The general observation is that ALE yields a poorer prediction performance than any other method. This is true both for the transfer and inline predictions. We also find that brain parcellations scores similar to the raw contrasts images, and closer to the inline predictions. We find that while the contrast-specific parcels and meta-analytic parcels methods do not return the same parcels, they produce very close results. We can thus use the full database to learn a single reference parcellation to perform meta-analysis.

Table 1. Prediction scores for inline and transfer learning. trans.= transfer; in.= inline; comp.= computation, sent.= sentences (reading), jabb.= jabberwocky; S= sentence with one word constituents, L= one constituent long sentence.

Names	Peaks		Contrasts		Parcels		Meta parcels	
	trans.	in.	trans.	in.	trans.	in.	trans.	in.
E1, comp./sent. → *E2*, comp./sent.	0.75	0.85	0.88	0.97	0.83	0.96	0.83	0.96
E2, comp./sent. → *E1*, comp./sent.	0.66	0.83	0.88	0.96	0.85	0.95	0.85	0.96
E3, jabb./French (L) → *E3*, jabb./French (S)	0.46	0.48	0.65	0.67	0.62	0.60	0.67	0.62
E3, jabb./French (S) → *E3*, jabb./French (L)	0.52	0.71	0.67	0.85	0.71	0.85	0.65	0.79
E3, jabb./French (L) → *E2*, Korean/French	0.65	0.46	0.73	0.79	0.65	0.81	0.76	0.85
E2, Korean/French → *E3*, jabb./French (L)	0.73	0.81	0.79	0.85	0.75	0.81	0.75	0.75

Fig. 1. Prediction performance relative to the best performing approach: inline prediction with raw contrasts images: the p-values indicate whether the associated methods are significantly poorer than the best performing method.
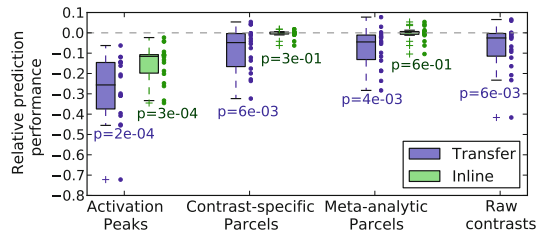
3.3 Experimental Results for Inference

Here we are interested in the *inference* problem: using transfer learning to help hypothesis testing on a target dataset. In the following, we only consider a specific transfer, namely the last line in Table 1: we learn a model discriminating French native speakers reading French or Korean, and apply it on another experiment in which French subjects had to read French or jabberwocky. This transfer is interesting as it involves two different experiments acquired on different scanners, and cognitive paradigms that share a similar expression, incomprehension of visual language stimuli. As can be seen in Table 1, the prediction scores of transfer learning as well as inline learning on this pair are acceptable although not excellent: French language and jabberwocky are difficult to separate.

Figure 2 gives the stability scores of the randomized logistic discriminating reading Korean from reading French for the different set of features –activation

peaks, raw contrasts, parcels learned on the training contrasts or on the full database. We can see that while learning at the voxel level or at the parcel level gives similar prediction performance (Table 1), the stability score maps are very different. At the voxel-level, with 70 subjects ($p = 40\,000$, $n = 70$) the recovery is limited to approximately 7 voxels without randomization: the recovery conditions are violated. As a result, the randomized logistic selects only the most predictive voxels. On the parcels, contrast-specific or meta-analytic (i.e., learned on the full database), the selection frequency highlights regions of the brain that are known to be relevant for language comprehension, including the left anterior superior temporal sulcus and the part of the temporal parietal junction (Wernicke's area).

We threshold the stability selection scores of the first experiment (Korean vs French) to select candidate voxels for the target experiment (jabberwocky vs French). As we want to perform a rough screening and would rather err on the side of false detections than false rejections, we take a very low threshold $\tau = .01$. Following our analysis above, the size of the largest group of correlated features that we can detect with such a threshold is on the order of $1/\tau \approx 100$. With 2000 parcels, this number corresponds to 5% of the brain, *i.e.* 8000 voxels, and we can safely consider that no fMRI contrasts is composed of groups of heavily correlated features larger than this fraction.

On the target experiment, we perform a standard group-level analysis with the voxels selected, testing for the difference between the two conditions, jabberwocky or French reading. We report results with p-values corrected for multiple comparisons at a given family-wise error rate (FWER) using Bonferroni correction, and for a given false discovery rate (FDR) using the Benjamini-Hochberg procedure. On table 2, we compare the number of detections and the detection rate, *i.e.* the fraction of voxels detected as significantly different, for a
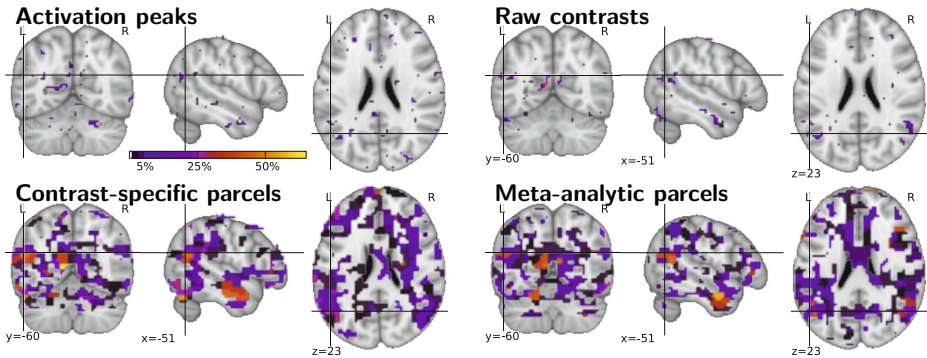
Fig. 2. Stability scores of the randomized logistic on the Korean versus French prediction of *E2* for the different set of features: the colormap represents the frequency at which a feature, parcel or voxel, was selected. The maps are thresholded at 1%.
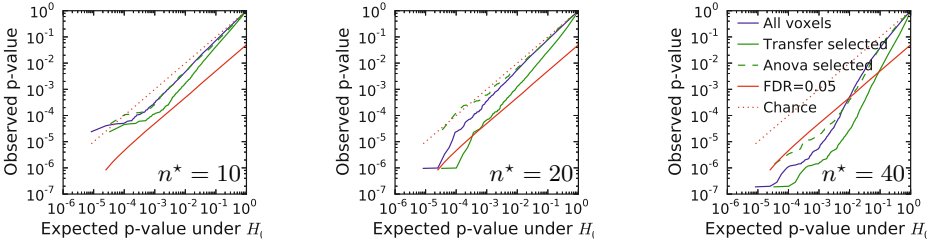
Fig. 3. Q-Q plots for the p-values with and without voxel selection by transfer learning, as well as FDR=0.05 threshold: **left** for a cohort size $n^\star = 10$, **middle** for a cohort size $n^\star = 20$, **right** for a cohort size $n^\star = 40$

Table 2. Number of detections at $p < 0.05$ for difference cohort size, for transfer learning and ANOVA. The percentage of detection is indicated in parenthesis.

$n^\star$	FWER corrected			FDR corrected		
	All voxels	Selection	ANOVA	All voxels	Selection	ANOVA
10	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
20	0 (0%)	3 (0.02%)	0 (0%)	0 (0%)	4 (0.027%)	0 (0%)
40	5 (0.0084%)	33 (0.22%)	2 (0.0014%)	143 (0.97%)	1339 (9%)	201 (1.4%)

full brain analysis and for an analysis limited to the voxel selection. We compare our voxel selection method to a one-way ANOVA, and find that transfer learning outperforms the ANOVA for all the cohort sizes. Figure 3 shows the Q-Q plots on which the Benjamini-Hochberg procedure is applied. We find that voxel selection by transfer learning improves both the absolute number of detections and the detection rate for FWER and FDR correction.

4   Conclusion

In this paper, we propose to improve the conditioning and power of statistical analyses in imaging studies, using a large meta-analytic database.

In a *transfer learning* scheme, we train on the database sparse discriminative models that are suited to the target experiment. Not only can the predictive power of these models can be useful to establish biomarkers, but also they perform feature selection that can increase the statistical power of a standard group analysis on new experiments, provided enough predictive features (voxels) can be recovered. Using brain parcellations, the discriminative model acts to screen parcels unlikely to be relevant in the target experiment, thus defining automatically ROIs.

Using a set of 3 fMRI studies related to language, we confirm experimentally that our transfer learning scheme is able to: i) perform accurate predictions on experiments acquired on a different scanner and with varying paradigm, ii) outperform the standard meta-analysis procedures based activation peaks, iii)

increase the statistical power in the target experiment by using the ROIs defined by the discriminative model.

In this work we manually select the contrast pairs since it is delicate to interpret a transfer learning score without good knowledge of the cognitive or clinical conditions under study. Future work will study automatic contrast pairs selection, *e.g.* by mining the descriptions of the experiments [4], to address the problem of synthesizing the ever-growing literature and data in medical research.

References

1. Sutton, A., Abrams, K., Jones, D., Sheldon, T., Song, F.: Methods for meta-analysis in medical research, West Sussex, UK, Chichester, England (2000)
2. Wager, T.D., Lindquist, M.A., Nichols, T.E., Kober, H., Snellenberg, J.X.V.: Evaluating the consistency and specificity of neuroimaging data using meta-analysis. Neuroimage 45(suppl. 1), 210–221 (2009)
3. Thirion, B., Pinel, P., Mériaux, S., Roche, A., Dehaene, S., Poline, J.B.: Analysis of a large fmri cohort: Statistical and methodological issues for group analyses. Neuroimage 35(1), 105–120 (2007)
4. Yarkoni, T., Poldrack, R.A., Nichols, T.E., Essen, D.C.V., Wager, T.D.: Large-scale automated synthesis of human functional neuroimaging data. Nat. Methods 8(8), 665–670 (2011)
5. Laird, A.R., Lancaster, J.L., Fox, P.T.: Brainmap: the social evolution of a human brain mapping database. Neuroinformatics 3(1), 65–78 (2005)
6. Salimi-Khorshidi, G., Smith, S.M., Keltner, J.R., Wager, T.D., Nichols, T.E.: Meta-analysis of neuroimaging data: a comparison of image-based and coordinate-based pooling of studies. Neuroimage 45(3), 810–823 (2009)
7. Pan, S., Yang, Q.: A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering 22(10), 1345–1359 (2010)
8. Bach, F.: Self-concordant analysis for logistic regression. Electronic Journal of Statistics 4, 384–414 (2010)
9. Meinshausen, N., Bühlmann, P.: Stability selection. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 72(4), 417–473 (2010)
10. Michel, V., Gramfort, A., Varoquaux, G., Eger, E., Keribin, C., Thirion, B.: A supervised clustering approach for fmri-based inference of brain states. Pattern Recognition (2011)
11. Pinel, P., Thirion, B., Meriaux, S., Jobert, A., Serres, J., Bihan, D.L., Poline, J.B., Dehaene, S.: Fast reproducible identification and large-scale databasing of individual functional cognitive networks. BMC Neurosci. 8, 91 (2007)
12. Pallier, C., Devauchelle, A.D., Dehaene, S.: Cortical representation of the constituent structure of sentences. Proc. Natl. Acad. Sci. U S A 108(6), 2522–2527 (2011)
13. Laird, A.R., Fox, P.M., Price, C.J., Glahn, D.C., Uecker, A.M., Lancaster, J.L., Turkeltaub, P.E., Kochunov, P., Fox, P.T.: Ale meta-analysis: controlling the false discovery rate and performing statistical contrasts. Hum. Brain Mapp. 25(1), 155–164 (2005)
14. Turkeltaub, P.E., Eden, G.F., Jones, K.M., Zeffiro, T.A.: Meta-analysis of the functional neuroanatomy of single-word reading: method and validation. Neuroimage 16(3 Pt 1), 765–780 (2002)

Radial Structure in the Preterm Cortex; Persistence of the Preterm Phenotype at Term Equivalent Age?

Andrew Melbourne[1], Giles S. Kendall[2], M. Jorge Cardoso[1], Roxanna Gunney[3], Nicola J. Robertson[2], Neil Marlow[2], and Sebastien Ourselin[1]

[1] Centre for Medical Image Computing, University College London, UK
[2] Academic Neonatology, EGA UCL Institute for Women's Health, London, UK
[3] Neuroradiology, Great Ormond Street Hospital for Children NHS Trust, London, UK

Abstract. Preterm birth increases the risk of perinatal brain injury and is believed to initiate a cascade of processes causing white matter damage resulting in subsequent neurological deficit; neonatal magnetic resonance imaging provides a number of potential biomarkers of this deficit. In this work we unify measures of the cortical folding pattern and of white matter integrity to establish correlation between grey and white matter derived properties. Diffusion weighted MRI has revealed that the cortical grey matter in the extremely preterm period exhibits a strong transient radial organisation suggesting neuronal axons are orientated towards the underlying white matter. This effect is lost during cortical maturation and is considered no longer visible on MRI at term equivalent age. Here we show that, in a group of 19 infants, radial organisation in the cortical grey matter remains detectable at term-equivalent age and that there is a strong anterior-posterior asymmetry. A group of three infants with moderate or severe abnormal white matter abnormality have significantly higher cortical grey matter radial organisation ($p < 0.02$), higher grey matter FA ($p < 0.01$) and a lower measure of cortical complexity ($p < 0.03$) than infants with normal or mild abnormal white matter abnormality; all measures associated with the preterm phenotype before term equivalent age. The novel combination of state-of-the-art imaging techniques, analysing grey-matter based spatial characteristics, may provide insight into the mechanism of neurodevelopmental deficits seen in infants with abnormal MR imaging at term equivalent age.

1 Introduction

Preterm brain injury is increasingly recognised to be an amalgamation of specific brain injury and abnormal brain development [1]. Advances in MR techniques including volumetric and cortical surface analysis [2] have begun to define differences in the brains of preterm infants imaged at term compared to term born controls and thus can help define biomarkers for prediction of subsequent neurodevelopment outcome. This preterm brain *phenotype* extends beyond reduction in volume and anisotropy of the cortical white matter to encompass decreased volume in the cortical grey matter, basal ganglia and cerebellum.

Very preterm birth coincides with a period of rapid brain growth and development [1]. During this period there is proliferation and migration of neurones from the subventricular zone to the cortex by radial migration. This results in radial cortical organisation; axons within the cortex lie perpendicular to the cortex, and this appears to be observed using diffusion tensor imaging as highly organised directional diffusion with consequent high fractional anisotropy. Examination of the principal eigenvectors suggest that axons are orientated perpendicular to the cortical surface towards the underlying white matter [3,4]. During normal cortical maturation the progressive development of multiple associative cortical interconnections results in a loss of this diffusion pattern by 35 weeks post menstrual age [3]. In addition, measures of gyrification are found to correlate with the underlying grey matter diffusion properties [5], although not after adjustment for gestational age at birth.

As preterm white matter injury is associated with a reduction in cortical grey matter volume we hypothesise it may also be associated with abnormal cortical development and organisation including higher cortical fractional anisotropy, lower white matter fractional anisotropy and the continued presence of cortical radial organisation visible on MRI at term equivalent age. In addition, measures of gyrification are found to correlate with the underlying grey matter diffusion properties [5]. We therefore also aim to correlate these findings to measures of cortical folding in babies with and without white matter injury on conventional MR imaging. Here, we develop a novel algorithm to specifically detect residual radial structure in the preterm cortex and show that its presence may be significantly higher in infants with moderate or severe white matter abnormality. The algorithm uses state-of-the-art image analysis techniques to develop, to the best of our knowledge, the first description of radial cortical architecture in very preterm infants at term equivalent age. Addressing this neonate-specific neuroscience question using such a technique may provide insight into the mechanism of neurodevelopmental deficits seen in infants with abnormal MR imaging at term equivalent age.

2   Methods

In this section we describe the steps needed to form the algorithm used for radial structure detection, comprising image segmentation, cortical surface extraction and diffusion parameter estimation and provide details of the 19 subject cohort of preterm infants scanned at term equivalent age.

2.1   Segmentation and Cortical Surface Analysis

Segmentation of the grey and white matter components of each infant is carried out using an adaptive neonatal-specific segmentation algorithm[1] [6]. The segmentation method estimates the cortical grey and white matter regions within an

---

[1] Available as part of the *niftyseg* package at: sourceforge.net/projects/niftyseg/

expectation maximisation routine whilst simultaneously spatially relaxing priors over the tissue classes. The algorithm incorporates a Markov random field for noise reduction, automated bias field correction and an implicit partial volume strategy. The interface between the segmented grey and white matter is found using a level-set technique, extracting a function at the surface upon which the gradient and curvature are implicitly defined at every position [7]. The boundary between grey and white matter is optimised using an evolving level set (1);

$$\frac{\delta\phi}{\delta t} + (\lambda_s s(\mathbf{x}, t) - \lambda_c c(\mathbf{x}, t))||\nabla\phi(\mathbf{x}, t)|| = 0 \qquad (1)$$

solving a partial differential equation for the function $\phi(\mathbf{x}, t)$ combining a signed distance function from the boundary, s, and a restriction on the curvature of the resulting function, c, with scalar weighting parameters $\lambda_s = 0.1$ and $\lambda_c = 0.4$ respectively. The level set is initialised along the white matter boundary and the result is a map of signed distances from the optimised GM/WM boundary.

We define measures of the gradient and curvature on this implicit surface, similar to the method used in [7]. The gradient of the level set is defined as perpendicular to the surface whilst the local Hessian matrix of second order derivatives can be use to define two curvatures; the eigenvalues of the local Hessian matrix define the principal curvatures, κ_1 and κ_2 (with $\kappa_1 > \kappa_2$) which summarise the local shape. Explicitly we define the signed shape index, S (2), describing how cup-like or saddle-like the surface is [8].

$$S = \frac{2}{\pi}\tan^{-1}\left(\frac{\kappa_2 + \kappa_1}{\kappa_1 - \kappa_2}\right) \qquad (2)$$

The distribution of the absolute values of shape index found on the surface may be summarised by the standard deviation of the absolute values of the shape index (implying that gyri and sulci are indistinguishable). Briefly, if the cortex has a complicated folding pattern, this is likely to introduce a larger range of possible shape-index values, and thus the standard deviation of this distribution, σ_S, will also be high. In contrast to global or slice-wise measures like the gyrification index, defined either automatically or manually [9,5], the shape index is defined at every position on the cortex, thus the standard deviation statistic may be defined over any arbitrary region.

2.2   Diffusion Imaging and Registration

Local fibre orientation in each voxel of the cortical grey matter is estimated by fitting the probabilistic ball and (two) stick model [10], thus each voxel contains a measure of its diffusivity and an estimate of the principal diffusion direction used to infer the underlying fibre orientation. Mean diffusivity images are registered to a single baseline and the diffusion information is then combined with the segmentation result by affine registration between the T1w image and the mean diffusivity image. Subsequently the GM, WM and level set segmentations are propagated to the diffusion space and we define the fractional anisotropy (FA) image in order to investigate spatial and group trends in the GM and WM FA.

2.3 Radial Structure Detection

After registration between the level set and the diffusion images, the radial direction perpendicular to the cortex in the diffusion space is defined by the gradient of the level set, g (Equation 3). If, at any position, the principal diffusion direction [10] in the grey matter, p, is parallel to the gradient of the level set function, the structure of the cortical grey matter can be considered radial (Equation 3). This measure of radial structure should not be confused with the radial component of the directional diffusivity given by the sum of the second and third eigenvalues of the diffusion tensor which has been shown to correlate with gestational age at birth [5]. Figure 1c illustrates the technique, relating the principal diffusion direction found using the ball and stick model to the direction across the underlying GM/WM boundary.

$$d = |\hat{g} \cdot \hat{p}| \qquad (3)$$

Integrating this function over the entire cortex results in a summary measure, D, of the radial structure for each infant, Equation 4 in spherical polar coordinates[2].

$$D = \int_{-\pi}^{\pi} \int_{0}^{\pi} \int_{0}^{\infty} |\hat{g} \cdot \hat{p}| r^2 sin\phi \delta r \delta \phi \delta \theta \qquad (4)$$

The measure, D, may also be subdivided into segments, $\delta\theta$ in the axial plane, θ ($D(\theta, \delta\theta)$), to observe any intra-subject spatial relationship along the medial lateral and anterior-posterior directions. This direction is expected to show the major trends in the folding pattern [9] and be sensitive to the known pattern of white matter myelination [11]. In any region of the cortex, $D(\theta, \delta\theta)$ may be compared with the underlying average grey and white matter FA and also to the standard deviation of absolute shape index values as described above, thus any observed radial structure can be correlated with known measures of cortical grey and white matter maturation. Visualisation of this information is provided using axially orientated rose plots; data is divided into segments and summary values found over each segment is presented by angle in the axial plane whilst the radius at any point represents the magnitude of the associated parameter. This style of visualisation allows variation in the observed parameters to be contrasted in the anterior-posterior and medial-lateral directions.

2.4 Data

Twenty infants born very preterm (less than 32 weeks completed gestation) underwent an MRI at term equivalent age. High-resolution T1-weighted imaging ($0.39 \times 0.39 \times 1mm$) and 2x30 directional diffusion tensor imaging ($b = 600 s.mm^{-2}$ at $0.9 \times 0.9 \times 3mm$) were acquired alongside two unweighted images

---

[2] Analytically, the expected value of the radial structure measure D over any arbitrary region containing an isotropic random distribution of grey matter diffusion directions is $D = 0.5$. Interestingly, this result is a 3D extension of the classic Buffon's needle problem.
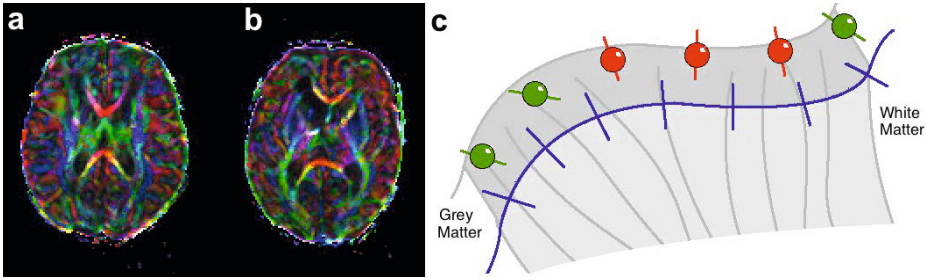
Fig. 1. Axial colour-coded FA slices for a) normal white matter appearance and b) moderate white matter abnormality visible on DTI. Note the suggestion of radially (perpendicular to the cortex) organised structure in the prefrontal cortex in b). c) illustrates the relationship between the principal diffusion direction found using the ball and stick model and the direction of the underlying GM/WM boundary.

and clinical T2 weighted images (3mm slices). One infant was subsequently removed from the study due to the presence of motion artefacts in the diffusion weighted sequence. Conventional MRI was assessed for WM abnormality using an established qualitative scoring system based on a grading system of 5 scales [12]. Images were scored by an experienced neuroradiologist blinded to clinical history and neurodevelopmental outcome. WM abnormality was classified as normal/mild or moderate/severe [12].

3   Results

Of the 19 babies analysed 11 were males, the average gestation was 25.8 weeks (range 22.9-30.7) with average birthweight of 822g (range 447-1185g). MR imaging was obtained at 40.8 weeks post menstrual age (range 38.3-44 weeks). Three babies were classified as moderate / severe white matter injury. The remaining 16 babies were classified as normal / mild using an established white matter scoring system [12].

Figure 1 shows colour-coded FA maps for four subjects, red colour coding corresponds to medial-lateral principal diffusion direction (PDD), green to anterior-posterior PDD and blue to superior-inferior PDD; pixel intensity represents the FA value. Figure 1a shows one infant with normal white matter appearance and little visible radial grey matter organisation. Figure 1b shows one infant with moderate white matter abnormality. The radial detection measure is calculated for the 16 infants with normal /mild white matter abnormality and for the 3 infants with moderate/severe white matter abnormality. Figure 2 illustrates the change in cortical grey and white matter properties with angle in the axial plane for the normal/mild group when divided into 64 equal angle segments (the anterior-posterior direction is aligned along the $90 - 270^o$ axis). Figure 2a measures the radial organisation (Equation 3) showing increased radiality in the frontal and pre-frontal region relative to the occipital region even at term equivalent age. In addition the pattern is broadly symmetric along the medial-lateral axis. Figure 2b and c illustrate the corresponding average cortical grey
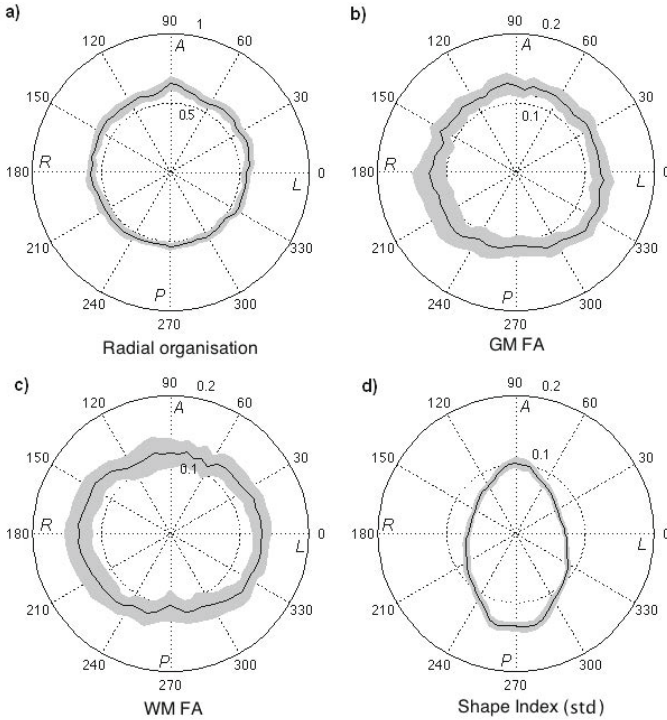
Fig. 2. Rose plots of each parameter with angle in the axial plane (normal/mild white matter abnormality group only): a) radial organisation measure; b) average GM FA value; c) average WM FA value (with each axial segment) and d) absolute shape index standard deviation. All plots show the mean in black, the grey region is $\pm\sigma$.

and underlying white matter FA values (calculated using the propagated segmentations) for each segment. The average grey matter FA is broadly isotropic around the cortex and appears higher in the pre-frontal region relative to the occipital region. The corresponding white matter pattern suggests higher FA along the medial-lateral axis, possibly representing the partial myelination of the corticospinal tracts. Total average grey and white matter FA are negatively correlated ($r = -0.56, p < 0.02$), which concords with the expected trajectories of white and grey matter maturation; the average FA associated particularly with the corticospinal tracts increases with gestation age whilst the observed grey matter FA reduces as a result of increased cortical inter-connectivity [5]. Figure 2d plots the standard deviation of the shape index values found in each region of the cortex, representing a measure of the cortical folding pattern. The cortical folding pattern is more complicated posterior than anterior. The total shape index standard deviation has a non-significant negative correlation with the grey matter FA ($r = -0.30, p = 0.22$).

Figure 3a and b show the spatial radial structure measure (Equation 3, $\mu \pm \sigma$) for the normal/mild and moderate/severe white matter abnormality groups
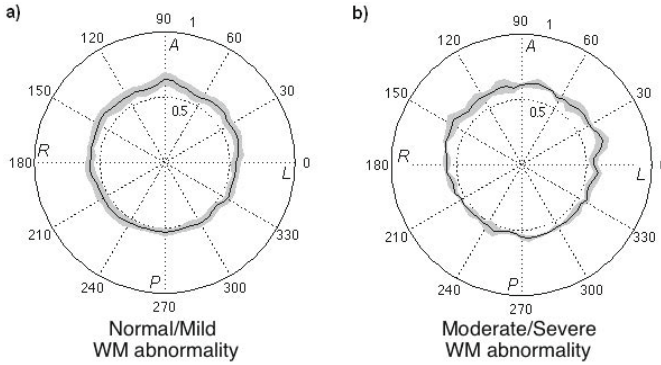
Fig. 3. Rose plots of radial structure for: a) normal or mild white matter abnormality (n=16) and b) moderate or severe white matter abnormality (n=3). All plots show the mean in black, the grey region is $\pm\sigma$. The value of D (Equation 4) is found to be significantly higher in the moderate/severe white matter abnormality group than the normal/mild group ($p < 0.02$), although the group sizes are disparate.

respectively. The moderate/severe group has a significantly higher measure of radial structure (D, Equation 4) than does the normal/mild group ($p < 0.02$), although the correlation is low. Additionally, the moderate/severe white matter abnormality group has significantly higher grey matter FA ($p < 0.01$) and significantly lower shape index standard deviation ($p < 0.03$) than infants with normal/mild white matter abnormality. Each of these features is associated with the preterm phenotype before term equivalent age, although it is not clear which is most informative.

4   Discussion

This work has used a combined analysis of segmentation and diffusion properties in both the grey and white matter to suggest a link between white matter abnormality on conventional MRI and structural organisation in very preterm infants scanned at term equivalent age. White and grey matter diffusion properties correlate with structural cortical folding information and retention of the preterm radial cortical architecture at term equivalent age in a small group of babies with moderate to severe white matter abnormality, particularly in the prefrontal cortex. The presence of this radial organisation in the cortex at term equivalent age may suggest that moderate to severe white matter abnormality is associated with a delay or disruption in normal cortical maturation, although we emphasise that the small group size makes general inference difficult. To our knowledge this is the first description of radial cortical architecture in very preterm infants at term equivalent age and may provide insight into the mechanism of neurodevelopmental deficits seen in infants with abnormal MR imaging at term equivalent age. Further work with serial MR acquisitions through the preterm period and beyond in combination with long term neurodevelopmental followup may help identify the nature of these changes and their association with

outcomes in childhood. Furthermore, formalisation of this technique into a robust analysis of grey-matter based spatial statistics may be a useful complement to existing white-matter techniques.

References

1. Volpe, J.J.: Brain injury in premature infants: a complex amalgam of destructive and developmental disturbances. Lancet Neurol. 8(1), 110–124 (2009)
2. Boardman, J.P., Craven, C., Valappil, S., Counsell, S.J., Dyet, L.E., Rueckert, D., Aljabar, P., Rutherford, M.A., Chew, A.T.M., Allsop, J.M., Cowan, F., Edwards, A.D.: A common neonatal image phenotype predicts adverse neurodevelopmental outcome in children born preterm. Neuroimage 52(2), 409–414 (2010)
3. McKinstry, R.C., Mathur, A., Miller, J.H., Ozcan, A., Snyder, A.Z., Schefft, G.L., Almli, C.R., Shiran, S.I., Conturo, T.E., Neil, J.J.: Radial organization of developing preterm human cerebral cortex revealed by non-invasive water diffusion anisotropy mri. Cereb. Cortex 12(12), 1237–1243 (2002)
4. Adams, E., Chau, V., Poskitt, K.J., Grunau, R.E., Synnes, A., Miller, S.P.: Tractography-based quantitation of corticospinal tract development in premature newborns. J. Pediatr. 156(6), 882–888, 888.e1 (2010)
5. Deipolyi, A.R., Mukherjee, P., Gill, K., Henry, R.G., Partridge, S.C., Veeraraghavan, S., Jin, H., Lu, Y., Miller, S.P., Ferriero, D.M., Vigneron, D.B., Barkovich, A.J.: Comparing microstructural and macrostructural development of the cerebral cortex in premature newborns: diffusion tensor imaging versus cortical gyration. Neuroimage 27(3), 579–586 (2005)
6. Cardoso, M.J., Melbourne, A., Kendall, G.S., Modat, M., Hagmann, C.F., Robertson, N.J., Marlow, N., Ourselin, S.: Adaptive Neonate Brain Segmentation. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part III. LNCS, vol. 6893, pp. 378–386. Springer, Heidelberg (2011)
7. Awate, S.P., Yushkevich, P.A., Song, Z., Licht, D.J., Gee, J.C.: Cerebral cortical folding analysis with multivariate modeling and testing: Studies on gender differences and neonatal development. Neuroimage 53(2), 450–459 (2010)
8. Koenderink, J.J., van Doorn, A.J.: Surface shape and curvature scales. Image and Vision Computing 10(8), 557–564 (1992)
9. Zilles, K., Armstrong, E., Schleicher, A., Kretschmann, H.J.: The human pattern of gyrification in the cerebral cortex. Anat. Embryol (Berl.) 179(2), 173–179 (1988)
10. Behrens, T.E.J., Woolrich, M.W., Jenkinson, M., Johansen-Berg, H., Nunes, R.G., Clare, S., Matthews, P.M., Brady, J.M., Smith, S.M.: Characterization and propagation of uncertainty in diffusion-weighted mr imaging. Magn. Reson. Med. 50(5), 1077–1088 (2003)
11. Brody, B.A., Kinney, H.C., Kloman, A.S., Gilles, F.H.: Sequence of central nervous system myelination in human infancy. i. an autopsy study of myelination. J. Neuropathol. Exp. Neurol. 46(3), 283–301 (1987)
12. Woodward, L.J., Anderson, P.J., Austin, N.C., Howard, K., Inder, T.E.: Neonatal mri to predict neurodevelopmental outcomes in preterm infants. N. Engl. J. Med. 355(7), 685–694 (2006)

Temporally-Constrained Group Sparse Learning for Longitudinal Data Analysis

Daoqiang Zhang[1,2], Jun Liu[3], and Dinggang Shen[1]

[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
[3] Imaging and Computer Vision Dept., Siemens Corporate Research, Princeton, NJ 08540
dqzhang@nuaa.edu.cn, junliu.nt@gmail.com, dgshen@med.unc.edu

Abstract. Sparse learning has recently received increasing attentions in neuroimaging research such as brain disease diagnosis and progression. Most existing studies focus on cross-sectional analysis, i.e., learning a sparse model based on single time-point of data. However, in some brain imaging applications, multiple time-points of data are often available, thus longitudinal analysis can be performed to better uncover the underlying disease progression patterns. In this paper, we propose a novel temporally-constrained group sparse learning method aiming for longitudinal analysis with multiple time-points of data. Specifically, for each time-point, we train a sparse linear regression model by using the imaging data and the corresponding responses, and further use the *group regularization* to group the weights corresponding to the same brain region across different time-points together. Moreover, to reflect the smooth changes between adjacent time-points of data, we also include two *smoothness regularization* terms into the objective function, i.e., one *fused smoothness* term which requires the differences between two successive weight vectors from adjacent time-points should be small, and another *output smoothness* term which requires the differences between outputs of two successive models from adjacent time-points should also be small. We develop an efficient algorithm to solve the new objective function with both group-sparsity and smoothness regularizations. We validate our method through estimation of clinical cognitive scores using imaging data at multiple time-points which are available in the Alzheimer's Disease Neuroimaging Initiative (ADNI) database.

1 Introduction

Neuroimaging plays an important role in characterizing the neurodegenerative process of many brain diseases such as Alzheimer's disease (AD). At present, a lot of pattern classification and regression methods have been developed for brain disease diagnosis and progression. Recently, sparse learning techniques have attracted more and more attentions due to their excellent performances in a series of neuroimaging applications on different modalities. For example, in a recent study [1], a voxel-based sparse classifier using L_1-norm regularized linear regression model, also known as the least absolute shrinkage and selection operator (LASSO) [2], was applied for classification

of AD and mild cognitive impairment (MCI) using magnetic resonance imaging (MRI) data, showing better performance than support vector machine (SVM) which is one of the state-of-the-art methods in brain imaging classification.

Following LASSO, several other advanced sparse learning models (i.e., LASSO variants) have also been recently used for solving problems in neuroimaging applications. For example, in [3], the elastic net which extends LASSO by imposing extra L_2-norm based regularizer to encourage a grouping effect, was recently used for identifying both neuroimaging and proteomic biomarkers for AD and MCI using MRI and proteomic data. In [4], a generalized sparse regularization with domain-specific knowledge was proposed for functional MRI (fMRI) based brain decoding. More recently, group LASSO [5], based on $L_{2,1}$-norm regularization, was used for jointly learning multiple tasks including both classification tasks (e.g., AD/MCI vs. healthy controls) and regression tasks (e.g., estimation of clinical cognitive scores) using MRI data in [6] and multimodal data including MRI, fluorodeoxyglucose positron emission tomography (FDG-PET) and cerebrospinal fluid (CSF) in [7], respectively. Here, the assumption of both methods is that multiple regression/classification variables are inherently related and essentially determined by the same underlying pathology, i.e., the diseased brain regions, and thus they can be solved together.

One commonplace of all above mentioned methods (i.e., LASSO and its variants) is that they aimed for cross-sectional analysis. In other words, only single-time-point imaging data (input) and single-time-point responses (output) are used for learning models in those methods. However, in some practical brain imaging applications, multiple-time-point data and/or multi-time-point responses are often available, thus longitudinal analysis can be performed to better uncover the underlying disease progression patterns [8]. According to the number of time-points in input and output of learning models, we can categorize them into the following four different learning problems: 1) Single-time-point Input and Single-time-point Output (SISO), 2) Single-time-point Input and Multi-time-points Output (SIMO), 3) Multi-time-points Input and Single-time-point Output (MISO), and 4) Multi-time-points Input and Multi-time-points Output (MIMO). Fig. 1 gives an illustration for these four different learning problems, with more detailed explanations given later in Section 2. To the best of our knowledge, most existing sparse models are aimed for the SISO problem (Fig. 1(a)), and it remains unknown in the literature on how to effectively use the longitudinal information in sparse learning to solve the other three problems (Fig. 1(b)-(d)).

In this paper, we address the above problems, i.e., SIMO, MISO and MIMO, which involves longitudinal information in either output or input or both. For this purpose, we develop a novel temporally-constrained group LASSO method, named as tgLASSO, which simultaneously includes the *group regularization* and the temporally *smoothness regularization* into its objective function. On one hand, as in group LASSO (gLASSO), for each time-point we train a sparse linear regression model by using the corresponding imaging data and responses at that time-point, and further use the *group regularization* to group the weights corresponding to the same brain region across different time points together. On the other hand, to reflect the smooth changes between adjacent time-points of data, we also introduce two smoothness regularization terms: 1) *fused smoothness* term which originates from fused LASSO [9] , for constraining the differences between two successive weight vectors from adjacent time-points to be small; 2) *output smoothness* term, for

constraining the differences between outputs of two successive models from adjacent time-points to be small. To the best of our knowledge, no previous sparse models ever use both the group-sparsity and the (fused plus output) smoothness regularizations into the objective function, for which we further develop a new efficient algorithm. We will use our proposed method for estimating clinical cognitive scores, e.g., Mini Mental State Examination (MMSE) and Alzheimer's Disease Assessment Scale - Cognitive Subscale (ADAS-Cog), by using MRI data from different time-points.
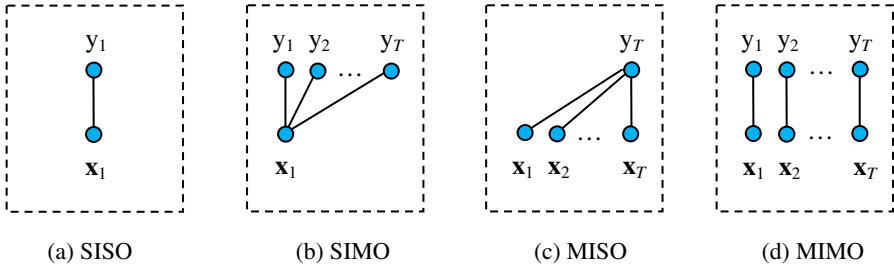
Fig. 1. Illustration on four different learning problems. Here, each edge represents a model, and the nodes $\mathbf{x}_j$ and y_j denote the imaging data (input) and clinical scores (output) at j-th time-point, respectively.

2   Method

In this section, we will introduce our temporally-constrained group LASSO (tgLASSO) method for longitudinal data analysis. We will first give our motivation and problem formulation in Section 2.1, followed by providing the objective function in Section 2.2 and the algorithmic solution in Section 2.3.

2.1   Motivation and Problem Formulation

Because of the neurodegenerative property of many brain diseases, e.g., AD and MCI, patients usually undergo a series of temporal changes reflected in MRI data and clinical scores (e.g., MMSE and ADAS-Cog for AD). Here, we want to estimate the clinical scores using MRI data. There are four different learning problems according to different number of time-points in both MRI data (input) and clinical scores (output), as shown in Fig. 1.

In the first learning problem, i.e., SISO as shown in Fig. 1(a), we want to estimate the clinical scores at a certain time-point, e.g., time-point 1 (baseline), by using imaging data from single time-point (e.g., baseline). Because both input and output are from single time-point, no longitudinal information is involved in this problem, and it can be easily solved by the existing sparse linear models, e.g., LASSO.

In the second learning problem, i.e., SIMO as shown in Fig. 1(b), we want to estimate the clinical scores at each time-point (ranging from 1 to T), by using imaging data from single time-point 1 (baseline). Similarly, in the third learning problem, i.e., MISO as shown in Fig. 1(c), we want to estimate the clinical scores at time-point T,

by using imaging data from all time-points (from 1 to T). Finally, in the fourth learning problem, i.e., MIMO as shown in Fig. 1(d), we want to estimate the clinical scores at each time-point j, by using imaging data from its corresponding time-point j, for $j = 1, \ldots, T$.

Unlike the first learning problem (SISO), the last three learning problems all involve longitudinal information, and thus cannot be directly solved using the existing sparse models. Also, it is worth noting that SIMO can be gotten from MIMO if setting $\mathbf{x}_j = \mathbf{x}_1$ (for $j = 1, \ldots, T$), and similarly MISO can be gotten from MIMO if setting $\mathbf{y}_j = \mathbf{y}_T$ (for $j = 1, \ldots, T$). For this reason, in this section we focus on MIMO and will further develop a new efficient algorithm to solve this new problem as below.

2.2   Objective Function

Assume that we have N training subjects, and each subject has T imaging data at T different time-points, represented as $\{\mathbf{x}_{i1}, \ldots, \mathbf{x}_{ij}, \ldots, \mathbf{x}_{iT}\}$, where $\mathbf{x}_{ij} \in \mathfrak{R}^{1 \times D}$ is a D-dimensional row vector. Denote $\mathbf{X}_j = [\mathbf{x}_{1j}; \ldots; \mathbf{x}_{ij}; \ldots; \mathbf{x}_{Nj}]$ ($\in \mathfrak{R}^{N \times D}$) and $\mathbf{y}_j$ ($\in \mathfrak{R}^{N \times 1}$) as the training data matrix (input) and the corresponding clinical scores at the j-th time-point, respectively. We use the linear model to estimate the clinical score from the imaging data $\mathbf{x}$ at the j-th time-point as $h_j(\mathbf{x}) = \mathbf{x}\mathbf{w}_j$, where the feature weight vector $\mathbf{w}_j \in \mathfrak{R}^D$. Let $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_j, \ldots, \mathbf{w}_T]$ ($\in \mathfrak{R}^{D \times T}$), then the objective function of our temporally-constrained group LASSO (tgLASSO) can be defined as follows

$$\min_{\mathbf{W}} \quad J(\mathbf{W}) = \frac{1}{2}\sum_{j=1}^{T}\left\|\mathbf{y}_j - \mathbf{X}_j\mathbf{w}_j\right\|_2^2 + R_g(\mathbf{W}) + R_s(\mathbf{W}) \tag{1}$$

Where $R_g(\mathbf{W})$ and $R_s(\mathbf{W})$ are the *group regularization* term and the *smoothness regularization* term, respectively, which are defined as below

$$R_g(\mathbf{W}) = \lambda_1\|\mathbf{W}\|_{2,1} = \lambda_1 \sum_{d=1}^{D}\|\mathbf{w}^d\|_2 \tag{2}$$

and

$$R_s(\mathbf{W}) = \lambda_2\sum_{j=1}^{T-1}\left\|\mathbf{w}_j - \mathbf{w}_{j+1}\right\|_1 + \lambda_3\sum_{j=1}^{T-1}\left\|\mathbf{X}_j\mathbf{w}_j - \mathbf{X}_{j+1}\mathbf{w}_{j+1}\right\|_2^2 \tag{3}$$

In Eq. 2, $\mathbf{w}^d$ is the d-th row vector of $\mathbf{W}$. It is worth noting that the use of L_2-norm on row vectors forces the weights corresponding to the d-th feature across multiple time-points to be grouped together and the further use of L_1-norm tends to select features based on the strength of T time-points jointly. The regularization parameter λ_1 controls the group sparsity of the linear models.

On the other hand, as shown in Eq. 3, the smoothness regularization consists of two parts. The first one as defined in the first term in Eq. 3 is called as the *fused smoothness* term which originates from fused LASSO [9], and its function is to constrain the differences between two successive weight vectors from adjacent time-points to be small. Also, it is worth noting that, due to the use of L_1-norm in the fused

smoothness term which encourages the sparsity on differences of weight vectors, there will be a lot of zeros in the components of the weigh difference vectors. In other words, a lot of components from adjacent weight vectors will be identical because of using the fused smoothness regularization. The second term in Eq. 3 is called as the *output smoothness* term which constrains the differences between outputs of two successive models from adjacent time-points to be small as well. The regularization parameters λ_2 and λ_3 balance the relative contributions of the two terms and also control the smoothness of the linear models. It is easy to know that when both λ_2 and λ_3 are zero, our method will reduce to group LASSO.

To the best of our knowledge, the objective function in Eq. 1 is the first time to simultaneously include both the group and the fused regularizations, which cannot be solved by the existing sparse models. Also, no previous studies consider using the output smoothness as extra regularizer. In the next section, we will develop a new efficient algorithm to solve the objective function in Eq. 1.

2.3   Efficient Iterative Solution

To minimize Eq. 1, we propose to use the iterative projected gradient descent approach [10]. Specifically, we separate the objective function in Eq. 1 to the smooth term

$$s(\mathbf{W}) = \frac{1}{2} \sum_{j=1}^{T} \left\| \mathbf{y}_j - \mathbf{X}_j \mathbf{w}_j \right\|_2^2 + \lambda_3 \sum_{j=1}^{T-1} \left\| \mathbf{X}_j \mathbf{w}_j - \mathbf{X}_{j+1} \mathbf{w}_{j+1} \right\|_2^2 \tag{4}$$

and the non-smooth term

$$n(\mathbf{W}) = \lambda_1 \|\mathbf{W}\|_{2,1} + \lambda_2 \sum_{j=1}^{T-1} \left\| \mathbf{w}_j - \mathbf{w}_{j+1} \right\|_1 \tag{5}$$

In each iteration k, the projected gradient descent contains two steps. Firstly, from $\mathbf{W}^{(k)}$, we compute

$$\mathbf{V}^{(k)} = \mathbf{W}^{(k)} - \gamma_k s'(\mathbf{W}^{(k)}) \tag{6}$$

where $s'(\mathbf{W}^{(k)})$ denotes the gradient of $s(\mathbf{W})$ at $\mathbf{W}^{(k)}$, and γ_k is the step size that can be determined by line search. Secondly, we set

$$\mathbf{W}^{(k+1)} = arg \min \frac{1}{2} \left\| \mathbf{W} - \mathbf{V}^{(k)} \right\|_2^2 + n(\mathbf{W}) \tag{7}$$

The problem in Eq. 7 is the proximal operator associated with the non-smooth term $n(\mathbf{W})$, and it can be computed by sequentially solving the proximal operator associated with the group Lasso penalty [5] and the proximal operator associated with the fuse Lasso penalty [9].

By utilizing the technique discussed in [10], the above projected gradient descent can be further accelerated to yield the accelerated gradient descent approach. Specifically, instead of performing gradient descent based on $\mathbf{W}^{(k)}$, we compute the search point

$$\mathbf{S}^{(k)} = \mathbf{W}^{(k)} + \alpha_k \left(\mathbf{W}^{(k)} - \mathbf{W}^{(k-1)} \right) \tag{8}$$

where α_k is a pre-defined variable [10], Then, we set

$$\mathbf{V}^{(k)} = \mathbf{S}^{(k)} - \gamma_i s'(\mathbf{S}^{(k)}) \tag{9}$$

Finally, we compute the new approximate solution as in Eq. 7. It can be shown that such a scheme can achieve a convergence rate of $O(1/l^2)$ for l iterations. For more details, please refer to [10].

3 Results

In this section, we validate our proposed tgLASSO method, with comparison to the existing LASSO and gLASSO methods, using 445 subjects (including 91 AD, 202 MCI, and 152 healthy controls) from the ADNI database. For each subject, there are MRI data as well as clinical scores including MMSE and ADAS-Cog, for the four different time-points, i.e., baseline, 6 months, 12 months, and 24 months which are denoted as T1, T2, T3 and T4, respectively. Our goal is to estimate the MMSE and ADAS-Cog scores at each of the four time-points using MRI data from corresponding time-point, which is a MIMO problem as shown in Fig. 1. It is worth noting that both SIMO and MISO problems can also be solved by our method as mentioned before. However, due to space limit, we do not report those results in this paper.

Standard image pre-processing is performed for all MRI images, including anterior commissure (AC) - posterior commissure (PC) correction, skull-stripping, removal of cerebellum, and segmentation of structural MR images into three different tissues: grey matter (GM), white matter (WM), and cerebrospinal fluid (CSF). Then, an atlas warping method [11] is used to register all different time-point images of each subject to a template with 93 manually labeled regions of interests (ROIs). For each of the 93 ROIs, we compute the GM tissue volume from the subject's MRI image as features.

In our experiments, 10-fold cross-validation is adopted to evaluate the performances of LASSO, gLASSO, and tgLASSO, by measuring the correlation coefficient between the actual clinical score and the estimated one. For all methods, the values of the parameters are determined by performing another cross-validation on the training data.

Fig. 2 shows the feature weight maps gotten from three different methods. Here, gLASSO and tgLASSO jointly learn the weight vectors for the four time-points, while LASSO learns each weight vector independently for each time-point. As can be seen from Fig. 2, due to the use of group regularization, gLASSO and tgLASSO obtain more grouped weights across different time-points than LASSO. Furthermore, due to the use of smoothness regularization, tgLASSO achieves more smooth weights across different time-points than other two methods. These properties are helpful to discover those intrinsic biomarkers relevant to brain diseases. For example, as shown in Fig. 2, among other disease related brain regions, both left and right hippocampal regions which are well-known AD-relevant biomarkers, are detected by tgLASSO, while only the left one can be detected by the other two methods.

On the other hand, Fig. 3 gives the comparisons of regression performances of the three methods in estimating MMSE and ADAS-Cog scores at four different time-points. As can be seen from Fig. 3, tgLASSO consistently outperforms the other two methods in estimating clinical scores for multiple time-points. In average, tgLASSO

achieves correlation coefficients of 0.613 and 0.639 for estimating MMSE and ADAS-Cog scores across all four time-points, respectively, while LASSO and gLASSO respectively achieve correlation coefficients of 0.569 and 0.587 for MMSE and 0.591 and 0.605 for ADAS-Cog. Fig. 3 also indicates that estimating later time-point scores often achieves better performance than estimating previous time-point scores. This may be because the relationship between imaging features and clinical scores becomes much stronger with progress of disease or brain aging, i.e., atrophy in the brain is more obvious in advanced disease and thus the related features are more distinctive and correlated to the clinical scores.
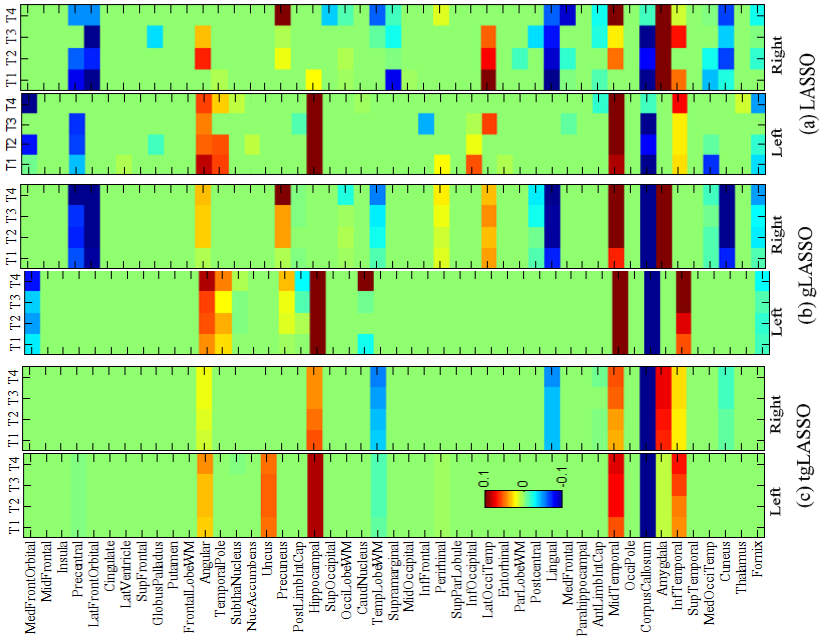
Fig. 2. Comparison of the feature weight maps of three different methods: (a) LASSO, (b) gLASSO, and (c) tgLASSO
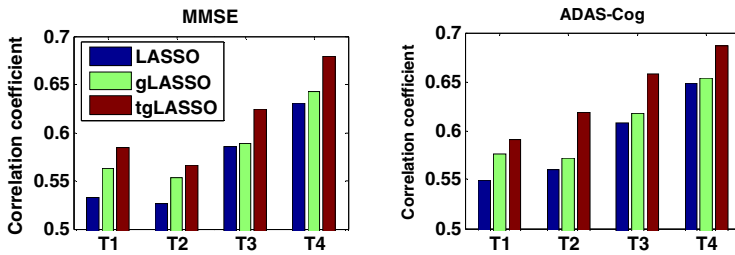
Fig. 3. Comparisons of regression performances of three different methods in estimating MMSE (left) and ADAS-Cog (right) scores

4   Conclusions

We have presented a new sparse learning method called tgLASSO for longitudinal data analysis with multiple time-points of data, which is different from most existing sparse learning methods focusing on cross-sectional analysis with single time-point of data. Our methodological contributions include: 1) proposing to simultaneously use group and (fused plus output) smoothness regularizations in sparse learning; 2) developing an efficient iterative algorithm for solving the new objective function. Experimental results on estimating clinical scores from imaging data at multiple time-points show the advantages of our method over the existing sparse methods on both regression performance and ability in discovering disease related imaging biomarkers.

References

1. Liu, M., Zhang, D., Shen, D.: Ensemble sparse classification of Alzheimer's disease. NeuroImage 60, 1106–1116 (2012)
2. Tibshirani, R.: Regression shrinkage and selection via the Lasso. J. Roy. Stat. Soc. B Met. 58, 267–288 (1996)
3. Shen, L., Kim, S., Qi, Y., Inlow, M., Swaminathan, S., Nho, K., Wan, J., Risacher, S.L., Shaw, L.M., Trojanowski, J.Q., Weiner, M.W., Saykin, A.J.: Identifying Neuroimaging and Proteomic Biomarkers for MCI and AD via the Elastic Net. In: Liu, T., Shen, D., Ibanez, L., Tao, X. (eds.) MBIA 2011. LNCS, vol. 7012, pp. 27–34. Springer, Heidelberg (2011)
4. Ng, B., Abugharbieh, R.: Generalized sparse regularization with application to fMRI brain decoding. Inf. Process Med. Imaging 22, 612–623 (2011)
5. Yuan, M., Lin, Y.: Model selection and estimation in regression with grouped variables. J. Roy Stat. Soc. B 68, 49–67 (2006)
6. Wang, H., Nie, F., Huang, H., Risacher, S., Saykin, A.J., Shen, L.: Identifying AD-Sensitive and Cognition-Relevant Imaging Biomarkers via Joint Classification and Regression. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part III. LNCS, vol. 6893, pp. 115–123. Springer, Heidelberg (2011)
7. Zhang, D., Shen, D.: Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease. NeuroImage 59, 895–907 (2012)
8. Xu, S., Styner, M., Gilmore, J., Piven, J., Gerig, G.: Multivariate nonlinear mixed model to analyze longitudinal image data: MRI study of early brain development. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, pp. 1–8 (2008)
9. Liu, J., Yuan, L., Ye, J.: An efficient algorithm for a class of fused lasso problems. In: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 323–332. ACM, Washington, DC (2010)
10. Beck, A., Teboulle, M.: A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems. SIAM J. Img. Sci. 2, 183–202 (2009)
11. Shen, D., Resnick, S.M., Davatzikos, C.: 4D HAMMER Image Registration Method for Longitudinal Study of Brain Changes. In: Proceedings of the Human Brain Mapping, New York City, USA (2003)

Feature Analysis for Parkinson's Disease Detection Based on Transcranial Sonography Image

Lei Chen[1,3], Johann Hagenah[2], and Alfred Mertins[1]

[1] Institute for Signal Processing, University of Luebeck, Germany
[2] Department of Neurology, University Hospital Schleswig-Holstein, Germany
[3] Graduate School, University of Luebeck, Germany

Abstract. Transcranial sonography (TCS) is a new tool for the diagnosis of Parkinson's disease (PD) according to a distinct hyperechogenic pattern in the substantia nigra (SN) region. However a procedure including rating scale of SN hyperechogenicity was required for a standard clinical setting with increased use. We applied the feature analysis method to a large TCS dataset that is relevant for clinical practice and includes the variability that is present under real conditions. In order to decrease the influence to the image properties from the different settings of ultrasound machine, we propose a local image analysis method using an invariant scale blob detection for the hyperechogenicity estimation. The local features are extracted from the detected blobs and the watershed regions in half of mesencephalon area. The performance of these features is evaluated by a feature-selection method. The cross validation results show that the local features could be used for PD detection.

Keywords: Parkinson's Disease, Transcranial Sonography, Blob detection, Feature analysis, local feature.

1 Introduction

Transcranial sonography (TCS) was used for the first time in a clinical study between a group of Parkinson's disease (PD) patients and healthy controls in 1995 [1]. For PD patients, the hyperechogenicity of the substantia nigra (SN) was significantly increased compared with controls. In 2002 the SN hyperechogenicity in PD was confirmed by another independent group [2]. By means of TCS, it is possible to determine the formation of idiopathic PD as well as monogenic forms of parkinsonism at an early state [3]. Furthermore, the SN area showed a distinct hyperechogenicity pattern on TCS for about 90% of PD patients, however, the structural abnormalities were not detected on CT and MRI scans [4]. These studies show that the SN hyperechogenicity is a valuable marker for PD diagnosis, especially for early diagnosis [5]. Compared to other clinical imaging modalities, the advantages of TCS include mobility, lack of side effects, and low cost. However, the quality of TCS images mainly dependents on the experience of the examiner and the acoustic bone window of the patient. With increased use,

a standardized procedure including rating scale of SN echogenicity was required for a standard clinical setting [7].

One solution to reduce investigator dependence of the diagnosis is to apply feature analysis to the image of the ipsilateral mesencephalon wing, which is close to the ultrasound probe as shown in Fig. 1. Firstly, the moment of inertia and Hu1-moment were calculated based on manually segmented half of mesencephalon (HoM) for separating control subjects from Parkin mutation carriers [5]. Then a hybrid feature extraction method which includes statistical, geometrical and texture features for the early PD risk assessment was proposed [8], which showed good performance of texture features (especially Gabor features). Thirdly, a texture analysis method that applied a bank of Gabor filters and gray-level co-occurrence matrices (GLCM) was used on TCS images [9]. After feature selection by sequential forward floating selection (SFFS), GLCM texture features were combined with Gabor features as a feature subset. The cross validation showed good results with the selected feature subset.
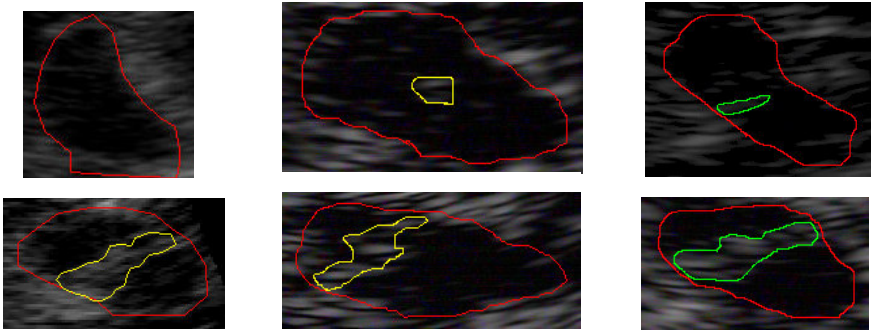
(a) Images from dataset1   (b) Images from dataset2   (c) Images from dataset3

Fig. 1. Manually segmented TCS images from Philips SONOS 5500. The first row is from healthy control subjects, and the second row from PD patients. The red marker indicates the upper HoM. Yellow/green markers show the SN area as a bright spot.

The last two previous works [8,9] analyzed data from only one ultrasound machine, and the selected features turned out to be sensitive to user settings and the ultrasound machine itself. In this paper, we collected three datasets that were acquired by different examiners with Philips SONOS 5500 in different periods. These datasets include the TCS images from PD and the healthy controls (HC). Actually, the properties of the TCS images, such as the gray values, the brightness, and the contrast, could be possibly affected by the different settings of the ultrasound machine used by different examiners. The mean and variance of the region of interest (HoM) in each TCS image were calculated and shown in Fig. 2. The variation between each dataset can be seen from TCS images in Fig. 1 and the statistical features of the images in Fig. 2. Our goal is to develop local features that are invariant to the illumination and contrast changes from

(a) 38 subjects of Parkinson's Disease   (b) 39 subjects of healthy control
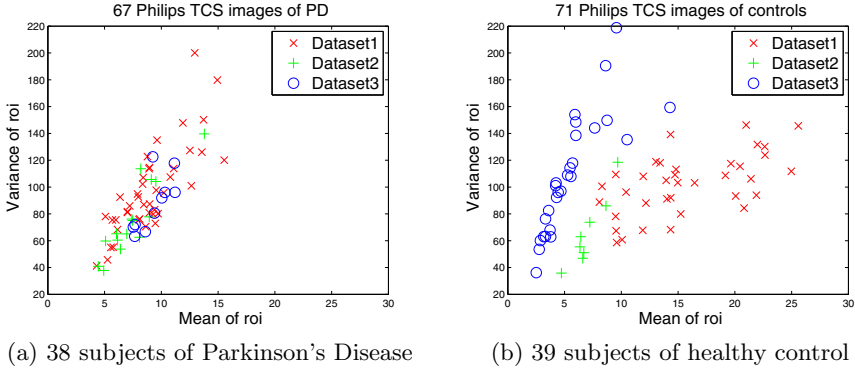
Fig. 2. The illustration about mean and variance of ROI (HoM) of 138 TCS images from Philips SONOS 5500

the different settings, even invariant to different ultrasound machines. The proposed local feature analysis method applies invariant blob detection to localize the hyperechogenicity area in HoM area and extracts local features based on watershed regions for the hyperechogenicity estimation.

2   Keypoint Localization

The hyperechogenicity of SN area consists of several bright spots in TCS image. The blob detection algorithm is stable under the monotonic changes in gray scale. The goal of this section is to localize the hyperechogenicity in HoM by the invariant scale blob detector. Based on space-scale theory, a multi-scale blob detector was proposed by Lindeberg [10], which could automatically select the appropriate scale for an observation. The scale space can be built using differential operators, such as Laplacian of Gaussian (LoG) and difference of Gaussians (DoG) filters. A brief framework for the invariant scale blob detector based on LoG is given by

$$\bigtriangledown^2_{\text{norm}} g = \sigma^2 \cdot \bigtriangledown^2 g(x, y; \sigma). \tag{1}$$

where σ is the standard deviation of the Gaussian $g(x, y; \sigma)$, and the scale-space representation $L(x, y; \sigma)$ of the image $f(x, y)$ is defined as

$$L(x, y; \sigma) = \bigtriangledown^2_{\text{norm}} g * f(x, y), \tag{2}$$

$$(\hat{x}, \hat{y}; \hat{\sigma}) = \arg[\text{extremum}_{(x, y; \sigma)} L(x, y; \sigma)], \tag{3}$$

where $(\hat{x}, \hat{y})$ corresponds to the center vector and $\hat{\sigma}$ to the scale vector of the detected blobs on each scale level. We suppose that one blob center $(\hat{x}_1, \hat{y}_1)$ is stable through the scale space, and a unique maximum over scales is given by

$$\partial_\sigma(L(\hat{x}_1, \hat{y}_1; \sigma)) = 0. \tag{4}$$

The evolution of blobs along scales was studied based on the idealized model patterns [10]. In practice, the amount of detected blobs on each scale level is different, and the centers of the same blobs might not be found at the same position on corresponding levels. One common solution is that a blob is detected if a local 3D extreme is present and its absolute value is higher than a threshold [11]. However, same blobs at different scales are not related and can be detected many times along the scale space. Our strategy is to link the trajectory of the same blobs along scale space and select the scale and center at the unique maxima that best represent each blob. For the presentation of this method, a phantom image was created as shown in Fig. 3(c). The linked pipelines for each detected blob from the phantom image are shown in Fig. 3(a). The corresponding local maxima of each pipe through scales are shown in Fig. 3(b). The final scale selection by (4) is shown in Fig. 3(c).
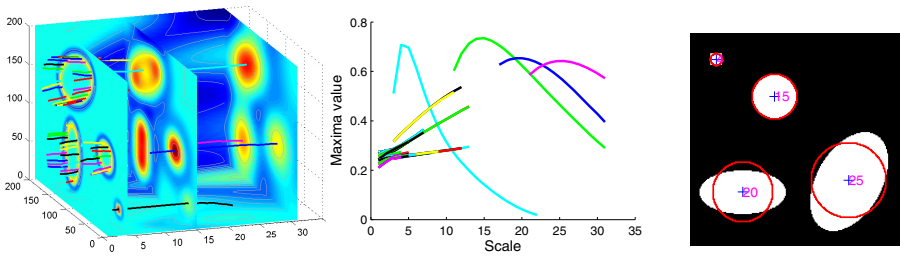
(a)Pipes through scale space   (b)Extrema of the pipes   (c)Scale selection result

Fig. 3. Blob scale selection from their trajectories along scale-space representation with LoG. (a) Three filtered images at scale σ = 4, 15, and 30. (b) Four global maxima at scales 4, 15, 20, and 25 were found from the connected trajectories. (c) Four corresponding blobs were detected and displayed on the phantom image.

In addition, the DoG is a close approximation to the scale normalized LoG, $\bigtriangledown^2_{norm}g$, given by

$$g(x, y; k\sigma) - g(x, y; \sigma) \approx (k-1) \bigtriangledown^2_{\mathrm{norm}} g, \qquad (5)$$

where the factor $(k-1)$ is constant over all scales and has almost no impact on the stability of extrema localization [12]. In this paper, DoG was applied for the construction of scale space. Actually, the analysis of scale-space maxima presents severe complications in TCS image, but the possible hyperechogenicity areas are localized by the proposed extrema selection method.

3   Local Feature Extraction

The mesencephalon is a butterfly-shape-like structure from the transverse view. The TCS image is obtained from the temporal acoustic bone window in a standardized axial mesencephalic imaging plane [5]. Only the HoM which is close to
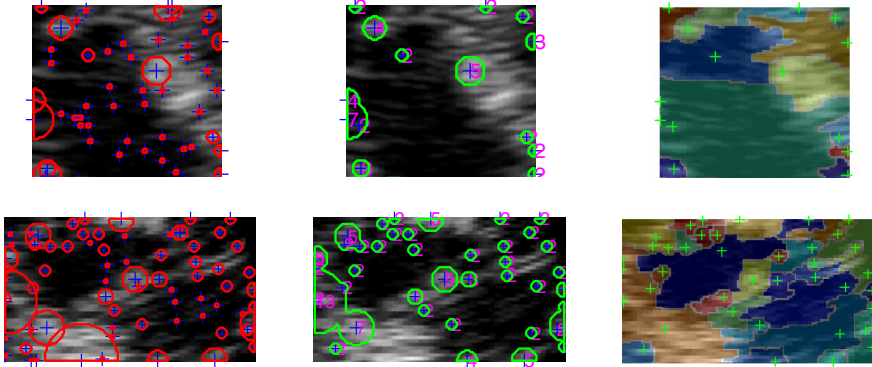
(a) Maxima in 26 neighbors   (b) Maxima in pilpeline   (c) Watershed regions

Fig. 4. (a) Detected maxima using DoG in 26 neighbors or (b) through pipe in scale space from control images (top row) and PD (bottom row) of Philips SONOS 5500. (c) Watershed segmentation results based on the detected blobs.

the probe is analyzed because of a decreased signal-to-noise ratio with increasing insonation depth. As a result, two TCS images from both sides are acquired per individual. It is better for this study not to include uncertainties that are attributed to the segmentation algorithm. Therefore HoM images were manually segmented by physicians and then analyzed for the estimation of the hyperechogenicity. The hyperechogenicity area is indicated with the blob detection as shown in Fig. 4. In the next step, a local image descriptor is needed to label the detected blob. The watershed algorithm [13] works on the gradient of an image, which is invariant to the brightness changes of the image. The watershed regions were thus segmented with the input of the detected blobs to estimate the hyperechogenicity in HoM.

Firstly, the blobs were detected with DoG operators in the HoM using the proposed extrema-selection method. The detection results of TCS images from Philips SONOS 5500 are shown in Fig. 4. The same blobs were prevented from being detected many times and the appropriate scales for each blob are indicated around the blob center as shown in Figs. 4(a) and (b). Secondly, based on the input of the detected blobs, the watershed regions were segmented and labeled by different color as shown in Fig. 4(c). Then, a selection procedure for the blob and watershed region was implemented with an ellipse mask filtering the false positives as shown in Fig. 5(a). From the prior knowledge of the anatomic location of SN, this mask is created from the ellipse which is fitted onto the ROI as mentioned in [5]. The values of the ellipse mask are calculated from their distance d to the minor ellipse axis. For $d < f$ (with f the distance between the focus point and the minor axis) the mask value is one. For $d \geq f$ the mask value is zero. Only the blobs that have big scale (For example, $\sigma \geq 3$) were taken into account as shown in Fig. 5(b). The watershed regions that are entirely within the ROI were considered as interesting areas. As a result, the selected blobs (indicated by green plus signs) and watershed regions are shown in Fig. 5(c).

(a) Ellipse mask   (b) Selected blobs   (c) Selected regions
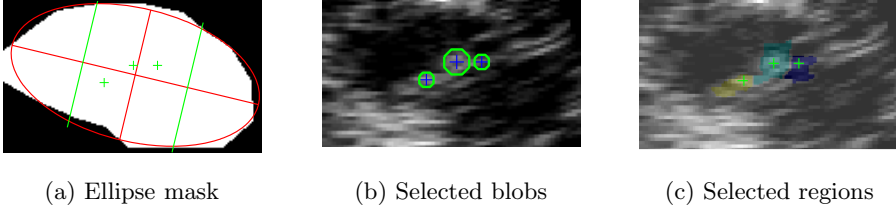
Fig. 5. (a) An ellipse is fitted with the ROI. Two green lines are parallel to the minor ellipse axis and across the two ellipse focuses, respectively. (b) The selected blobs (green sign) and (c) the selected watershed regions which are inside of the ellipse mask.

For the estimation of the hyperechogenicity, nine local features $F1...F9$ were extracted based on the selected blobs and watershed regions in HoM. Entropy is used to measure the randomness of a local region. The parameters shape and scale of a Weibull approximation [14] of the gradient distribution were determined by maximum likelihood estimation [13] and used as local image features. The calculation of entropy and the estimation of Weibull distribution parameters were obtained from the gradient images after Gaussian smoothing. Considering the image scaling, the features F1 and F2 were normalized by the corresponding HoM area. The local features are shown as follows:

F1,F3: Area and entropy of all selected watershed regions
F2,F4: Area and entropy of all selected blobs
F5,F6: Weibull parameters (a,b) of all selected watershed regions and blobs
F7: The scale of the biggest detected blob
F8,F9: Entropy of the biggest blob and HoM

4 Experimental Results

The experiments were based on three data sets which were obtained with Philips SONOS 5500 by different examiners. Dataset 1 includes 42 TCS images from 23 PD patients and 36 TCS images from 21 healthy controls. Dataset 2 includes 15 PD TCS images from ten PD patients and eight control images from four controls. The last dataset consisted of ten PD TCS images from five PD patients and 27 TCS from 14 controls. Totally, this large dataset includes 67 PD images from 38 PD patients and 71 control images from 39 healthy subjects.

The outline of the framework is as follows: First, the dataset is classified using the selected feature subsets F(17, 25, 26, 27, 29) from [8] and F(17, 77) from [9]. Secondly, based on the manually segmented HoM images which were marked by the physicians, the suspicious hyperechogenicity areas were localized by the invariant scale blob detection method. Then, the watershed-segmentation algorithm was applied to the gradient image after Gaussian smoothing. At last, local features were extracted based on the selected blobs and the watershed regions. These local features were evaluated by the feature-selection method SFFS. The criterion function of SFFS was the accuracy of the SVM classifier.

The training of SVMs was carried out with sequential minimal optimization (SMO) and a linear kernel. The SVM classification results were cross validated with the leave-one-out method.

The feature analysis results are shown in Table 1. Based on this dataset, the features found in [8] and [9] achieved 76.81% and 48.55% correct rate, respectively. Five local features $F(3,7,8,1,9)$ were selected with SFFS based on this dataset. Using the selected local featrues, the classification accuracy reached 72.46%, which was better than the Gabor feature and GLCM feature from [9]. To test how the feature sets perform when standard operations such as brightness and contrast normalization are carried out, for each image the intensity values in the ROI were normalized to the range $[0, 255]$. The results in the right column of Table 1 show that the local features are invariant to illumination changes from the image normalization and outperform the other features under such conditions.

In another experiment, an SVM classifier was used to evaluate the performance of the three selected feature subsets when the training is carried out on other datasets than the test. We used Datasets 1 and 3 for Training and Dataset 2 for test. The classification results are listed in Table 2. They show that the classifier with the selected local features works better than the others when training and test conditions are different.

Table 1. Feature analysis and SVMs cross-validation results on the large dataset

Dataset 1,2,3	Accuracy	Confusion matrix	Accuracy(normalized data)
$F(17, 25, 26, 27, 29)$ from [8]	76.81%	$\begin{pmatrix} 63 & 4 \\ 28 & 43 \end{pmatrix}$	71.01%
$F(17, 77)$ from [9]	48.55%	$\begin{pmatrix} 40 & 27 \\ 44 & 27 \end{pmatrix}$	58.70%
Local feature $F(3, 7, 8, 1, 9)$	72.46%	$\begin{pmatrix} 52 & 15 \\ 23 & 48 \end{pmatrix}$	72.46%

Table 2. Classification results of the three selected feature subsets

Training data(Dataset 1,3), test data (Dataset2)	Accuracy	Confusion matrix
$F(17, 25, 26, 27, 29)$ from [8]	65.22%	$\begin{pmatrix} 15 & 0 \\ 8 & 0 \end{pmatrix}$
$F(17, 77)$ from [9]	60.87%	$\begin{pmatrix} 14 & 1 \\ 8 & 0 \end{pmatrix}$
Local feature $F(3, 7, 8, 1, 9)$	78.26%	$\begin{pmatrix} 14 & 1 \\ 4 & 4 \end{pmatrix}$

5   Conclusions

We have analyzed the selected features from two previous works and nine new local features based on a large dataset of TCS images. In particular, the local features are invariant to the monotonic changes in gray scale. Almost all possible locations of hyperechogenicity in HoM area could be indicated by the proposed

invariant scale blob detection. Moreover, the watershed segmentation was applied to segment the ROI for PD detection. Of course, the current results depend on the manual segmentation of HoM area by physician. An automatic segmentation algorithm could be implemented for localization of the HoM area. Even though the appearance of mesencephalon can vary considerably across subjects, the prior knowledge of anatomic shape and location of SN can be utilized for the improvement of the selection strategy. The keypoint detection would be improved with shape estimation, and more robust and precise local image descriptors of hyperechogenicity may be developed for PD detection.

References

1. Becker, G., Seufert, J., Bogdahn, U., Reichmann, H., Reiners, K.: Degeneration of substantia nigra in chronic Parkinson's diseasvisualized by transcranial color-coded real-time sonograph. Neurology 45, 182–184 (1995)
2. Walter, U., Wittstock, M., Benecke, R., Dressler, D.: Substantia nigra echogenicity is normal in non-extrapyramidal cerebral disorders but increased in Parkinson's disease. J. Neural Transm. 109, 191–196 (2002)
3. Behnke, S., Berg, D., Becker, G.: Does ultrasound disclose a vulnerability factor for Parkinson's disease? J. Neural 250(suppl.1), I24–I27 (2003)
4. Hagenah, J.M., Hedrich, K., Becker, B., Pramstaller, P.P., Seidel, G., Klein, C.: Distinguishing early-onset PD from dopa-responsive dystonia with transcranial sonography. Neurology 66, 1951–1952 (2006)
5. Kier, C., Seidel, G., Bregemann, N., Hagenah, J., Klein, C., Aach, T., Mertins, A.: Transcranial Sonography as Early Indicator for Genetic Parkinson's Disease. In: Vander Sloten, J., Verdonck, P., Nyssen, M., Haueisen, J. (eds.) ECIFMBE 2008. IFMBE Proceedings, pp. 456–459. Springer, Heidelberg (2009)
6. Kier, C., Cyrus, C., Seidel, G., Hofmann, U., 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, S83 (2007)
7. Berg, D., Steinberger, J.D., Warren Olanow, C., Naidich, T.P., Yousry, T.A.: Milestones in Magnetic Resonance Imaging and Transcranial Sonography of Movement Disorders. Movement Disorders 26 (2011)
8. Chen, L., Seidel, G., Mertins, A.: Multiple Feature Extraction for Early Parkinson Risk Assessment Based on Transcranial Sonography Image. In: Proceedings of 2010 IEEE 17th International Conference on Image Processing (2010)
9. Chen, L., Hagenah, J.M., Mertins, A.: Texture Analysis Using Gabor filter Based on Transcranial Sonography Image. In: Bildverarbeitung Für die Medizin 2011, pp. 249–253 (2011)
10. Lindeberg, T.: Feature Detection with Automatic Scale Selection. International Journal of Computer Vision 30(2) (1998)
11. Ferraz, L., Binefa, X.: A Scale Invariant Interest Point Detector for Discriminative Blob Detection. In: Araujo, H., Mendonça, A.M., Pinho, A.J., Torres, M.I. (eds.) IbPRIA 2009. LNCS, vol. 5524, pp. 233–240. Springer, Heidelberg (2009)
12. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60, 91–110 (2004)
13. Meyer, F.: Topographic distance and watershed lines. Signal Processing 38, 113–125 (1994)
14. Devroye, L.: Non-Uniform random variate generation. Springer, New York (1986)

Longitudinal Image Registration with Non-uniform Appearance Change

Istvan Csapo[1], Brad Davis[3], Yundi Shi[1], Mar Sanchez[4], Martin Styner[1], and Marc Niethammer[1,2]

[1] University of North Carolina at Chapel Hill, NC
[2] Biomedical Research Imaging Center, UNC Chapel Hill, NC
[3] Kitware, Inc., Carrboro, NC
[4] Emory University, Atlanta, GA
icsapo@cs.unc.edu

Abstract. Longitudinal imaging studies are frequently used to investigate temporal changes in brain morphology. Image intensity may also change over time, for example when studying brain maturation. However, such intensity changes are not accounted for in image similarity measures for standard image registration methods. Hence, (i) local similarity measures, (ii) methods estimating intensity transformations between images, and (iii) metamorphosis approaches have been developed to either achieve robustness with respect to intensity changes or to simultaneously capture spatial and intensity changes. For these methods, longitudinal intensity changes are not explicitly modeled and images are treated as independent static samples. Here, we propose a *model-based image similarity measure* for longitudinal image registration in the presence of spatially non-uniform intensity change.

1 Introduction

To study changes that occur during brain development, neurodegeneration, or disease progression in general, longitudinal imaging studies are important. Spatial correspondences almost always need to be established between images for longitudinal analysis through image registration. Most image registration methods have been developed to

Fig. 1. Brain slices and magnifications for a monkey at ages 2 weeks, 3, 6, 12, and 18 months. White matter appearance changes locally as axons are myelinated during brain development.

align images that are similar in appearance or structure. If such similarity is not given (e.g., in case of pathologies or for pre- and post-surgery images) cost function masking is typically used to discard image regions without correspondence from the registration. Such strict exclusion is not always desirable. When investigating brain maturation for example (our target application in this paper) valid correspondences for the complete brain are expected to exist. However,

brain appearance changes continuously over time due to biological tissue changes (here, myelination of white matter [1,12]) and adversely affects image registration results [9].

The effect of appearance change on the result of an image registration depends on the chosen transformation model and the chosen image similarity measure. Generally, transformation models with few degrees of freedom (such as rigid or affine transformations) are affected less by local changes in image appearance than transformation models which can capture localized spatial changes, such as elastic or fluid models. In particular, we have previously shown that affine methods perform well even in the presence of strong non-uniform appearance change, while deformable methods introduce erroneous local deformations in order to resolve inconsistencies in appearance [3]. However, transformation models which can capture local deformations are desirable for many longitudinal studies as changes in morphology tends to be spatially non-uniform.

For longitudinal registration, temporal regularization of the transformation model has been explored recently. This is motivated by the assumption that unrealistic local changes can be avoided by enforcing temporal smoothness of a transformation [5,7]. In this paper we instead focus on the complementary problem of determining an appropriate image similarity measure for longitudinal registration in the presence of temporal changes in image intensity.

Approaches which address non-uniform intensity changes have mainly addressed registration for image-pairs so far and either rely on local image uniformities [9,13] or try to estimate image appearance changes jointly with an image transform [8,11,10]. Often (e.g., for bias field compensation in magnetic resonance imaging), image intensity changes are assumed to be smooth. Our proposed approach in contrast, estimates local longitudinal models of intensity change using all available images. Our approach alternates between parameter estimation for the local models of intensity change and estimation of the spatial transformation. Image similarities are computed relative to the estimated intensity models, hence accounting for local changes in image intensities.

Section 2 introduces the model-based image similarity measure (mSM). Section 3 discusses parameter estimation. Section 4 describes the performed experiments and discusses results. The paper concludes with a summary and outlook on future work.

2   Model-Based Similarity Measure

Assume we have an image intensity model $\hat{I}(x, t; p)$ which for a parameterization, p, describes the expected intensity values for a given point x at a time t. This model is defined in a spatially fixed target image. Then, instead of registering a measured image I_i at t_i to a fixed target image I_T we can register it to its corresponding intensity-adjusted image $\hat{I}(x, t_i; p)$, effectively removing temporal intensity changes for a good model and a good parameterization, p. Hence,

$$\text{Sim}(I_i \circ \Phi_i, I_T) \quad \text{is replaced by} \quad \text{Sim}(I_i \circ \Phi_i, \hat{I}(x, t_i; p)),$$

where $\text{Sim}(\cdot, \cdot)$ is any chosen similarity measure (e.g., sum of squared differences (SSD), normalized cross correlation, or mutual information), and Φ_i is the map from image I_i to the spatially fixed target space. Since our method aims to create an intensity adjusted model $\hat{I}$ that matches the appearance of the source image, we use SSD in this paper. We call the intensity-adjusted SSD similarity measure a sum of squared residual (SSR) model, where the residual is defined as the difference between the predicted and the measured intensity value.

2.1 General Local Intensity Model Estimation for SSD

Since SSR is a local image similarity measure, for a given set of N measurement images $\{I_i\}$ at times $\{t_i\}$ we can write the full longitudinal similarity measure as the sum over the individual SSRs, i.e.,

$$SSR(\{I_i\}; p) = \sum_{i=0}^{N-1} \int_{\Omega} (I_i \circ \Phi_i(x) - \hat{I}(x, t_i; p))^2 \, dx,$$

where Ω is the image domain of the fixed image. For given spatial transforms Φ_i this is simply a least-squares parameter estimation problem given the measurements $\{I_i \circ \Phi_i(x)\}$ and the predicted model values $\{\hat{I}(x, t_i; p)\}$. We use alternating optimization with respect to the intensity model parameters, p, and the spatial transformations Φ_i to convergence (see Sec. 3).

2.2 Logistic Intensity Model with Elastic Deformation

SSR can be combined with any model for intensity change, ranging from a given constant target image (the trivial model), linear models, and splines to models more closely adapted to the myelination process we are interested in capturing during neurodevelopment. Since the myelination process exhibits a rapid increase during early brain development followed by a gradual leveling off [4], nonlinear appearance models are justified. In this paper we investigate the logistic model
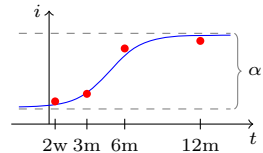
Fig. 2. Logistic intensity model

$$\hat{I}(x, t; \alpha(x), \beta(x), k(x)) = \frac{\alpha(x)}{1 + \beta(x)e^{-k(x)t}} \, , \tag{1}$$

which is often used in growth studies [6]. Here, α, β, and k are spatially varying model parameters with biological meaning, k being the maximum rate of intensity change, α the maximum increase of white matter intensity during myelination, and β is related to the onset time of myelination (see Fig. 2). Assuming that both unmyelinated and fully myelinated white matter intensities are spatially uniform we keep α constant as the difference between myelinated (upper asymptote) and unmyelinated (lower asymptote) white matter intensities. This is a simplifying, but reasonable assumption since intensity inhomogeneities in

unmyelinated or myelinated white matter are small compared to the white matter intensity change due to the myelination process itself [1].

3   Parameter Estimation

Once the parameters for the local intensity models are known, SSR can be used to replace the image similarity measure in any longitudinal registration method. Here, we use an elastic deformation model (Sec. 3.1) and jointly estimate the parameters for the intensity model (Sec. 3.2).

3.1   Registration Model

The growth process of the brain not only includes appearance change but complex morphological changes as well, hence the need for a deformable transformation model. To single out plausible deformations, we use (for simplicity) an elastic regularizer [2] defined on the displacement field u as

$$S[u] = \int_{\Omega} \frac{\mu}{4} \sum_{j,k=1}^{d} \underbrace{(\partial_{x_j} u_k + \partial_{x_k} u_j)^2}_{\text{rigidity}} + \frac{\lambda}{2} \underbrace{(\text{div } u)^2}_{\text{volume change}} \, dx,$$

where μ (=1) and λ (=0) are the Lamé constants that control elastic behavior, and the div is the divergence operator defined as $\nabla \cdot u$, where ∇ is the gradient operator. Registrations over time then decouple into pairwise registration between the intensity-adjusted target image and a given source image I_i. This is sufficient for our test of the longitudinal image similarity measure, but could easily be combined with a spatio-temporal regularizer which would then directly couple the transformations between the images of a time-series (instead of only having an indirect coupling through the model-based similarity measure).

3.2   Model Parameter Estimation

We estimate the intensity model parameters only within the white matter (segmentation was obtained at the last time-point with an atlas-based segmentation method and propagated to earlier time-points [14]) where image appearance changes non-uniformly over time; for simplicity, gray matter intensity was assumed to stay constant.

Note that overestimating the white matter results in fitting the model to the surrounding gray matter voxels. This has negligible effect on the registration, since the model can capture the constant gray matter intensities. Underestimating the white matter, on the other hand, can lead to uncorrected intensities near the white-gray matter boundary and introduce erroneous local deformations.

Instead of estimating the parameters independently for each voxel, spatial regularization was achieved by estimating the parameters from overlapping local $3 \times 3 \times 3$ neighborhoods using robust statistics (median of the parameters).

The algorithm is defined as follows

0) Initialize model $\hat{I}$ parameters to $p = p_0$.
1) Affinely pre-register images $\{I_i\}$ to $\hat{I}$.
2) Estimate the appearance of $\hat{I}$ at times $\{t_i\}$, giving $\{\hat{I}(t_i)\}$.
3) Estimate displacement fields $\{u_i\}$ by registering images $\{I_i\}$ to $\{\hat{I}(t_i)\}$.
4) Estimate model parameters p from the registered images $\{I_i \circ u_i\}$.
5) Repeat from step 2 until convergence.

The algorithm terminates once the registration energy decreases by less than a given tolerance between subsequent iterations. In all our experiments only few iterations (typically less than 5) were required. A more in-depth numerical convergence analysis should be part of future work. If desired, a prior model defined in the target image (a form of intensity model parameter atlas) could easily be integrated into this framework.

4   Experimental Results

We compared the model-based similarity measure to mutual information (MI) on sets of longitudinal magnetic resonance

12 mo   6 mo   3 mo   2 wk
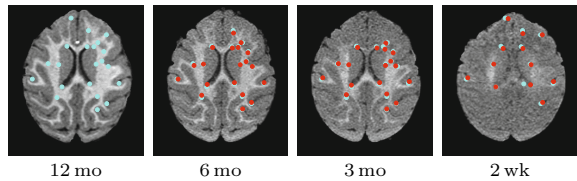
Fig. 3. Corresponding target (cyan) and source (red) landmarks for a single subject

images of 9 monkeys, each with 4 time-points. Each set was affinely pre-registered and intensity normalized so that the gray matter intensity distributions matched after normalization (gray matter intensity generally stays constant over time).

We registered 3D images of the three early time-points $I_{2\mathrm{wk}}, I_{3\mathrm{mo}}, I_{6\mathrm{mo}}$ to the target image $I_{12\mathrm{mo}}$ with an elastic registration method. Since the ground truth deformations were not known, manually selected landmarks (Fig. 3) identified corresponding regions of the brain at the different time-points (10-20 landmarks in a single slice for each of the 4 time-points in all 9 subjects; the landmarks were picked based on geometric considerations). The distance between transformed and target landmarks yielded registration accuracy. Figure 4 shows the experimental setup.

Note that for the model-based method the target image is not $I_{12\mathrm{mo}}$ but the model $\hat{I}_{12\mathrm{mo}}$ estimated at time t of the source image. Here, we estimate the intensity change of $I_{12\mathrm{mo}}$ as white matter segmentation is easily obtained given the good gray matter white matter contrast, but other time-points could be used.

With MI, the registration method accounts for both the non-uniform white matter appearance change and the morphological changes due to growth through large local deformations. This is especially apparent for registrations between $I_{2\mathrm{wk}}$ (and to a lesser extent $I_{3\mathrm{mo}}$) and the target $I_{12\mathrm{mo}}$ and suggests large
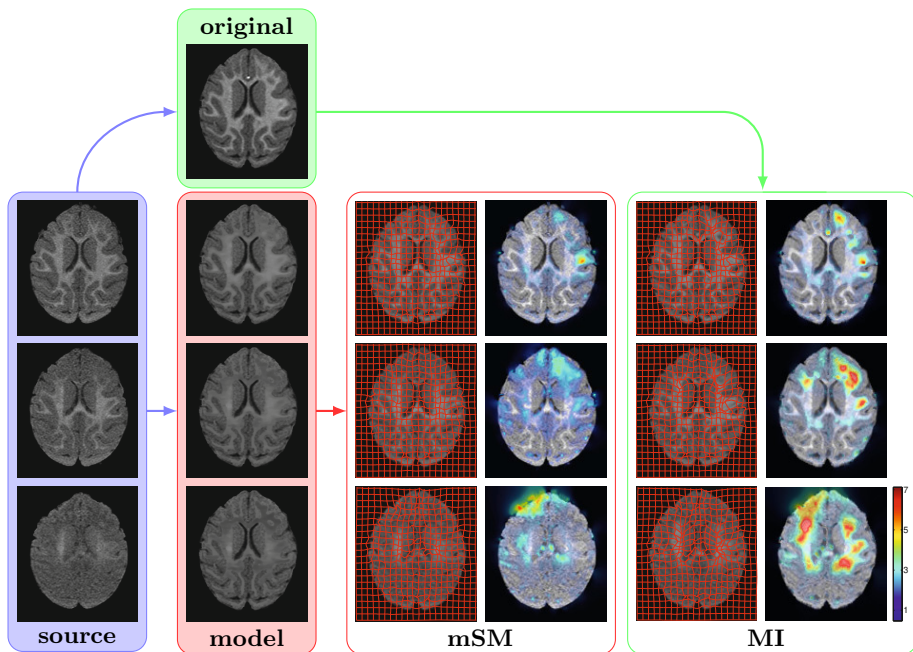
Fig. 4. Experimental setup and results for a single subject. To test MI, the source images (blue; from bottom: I_{2wk}, I_{3mo}, I_{6mo}) are registered to the latest time-point I_{12mo} (green). The resulting deformation field and the magnitude of the deformations (in pixels) is shown in the right panel. For mSM, the source images are registered to the model (red) that estimates the appearance of I_{12mo} at the corresponding time of each source image (results in middle panel).

local morphological changes contradictory to normal brain development [4]. The landmark mismatch results (Fig. 5) show that both mutual information and the model-based approach perform well in the absence of large intensity non-uniformity, however, mSM consistently introduces smaller erroneous deformations than MI.

Table 1 shows the aggregate results of the landmark mismatch calculations for both methods. The model-based approach can account for appearance change by adjusting the intensity of the model image (see the estimated model images in Fig. 4) and therefore is most beneficial when the change in appearance between the source and target image is large (I_{2wk}, I_{3mo}).

We also compared 1st and 2nd degree polynomial intensity models to the logistic model and found no significant difference. This is expected with only 4 time-points as even the 1st degree model can reasonably estimate the local appearance changes. However, we expect the logistic model to outperform the simpler models in larger studies with more time-points or when the model appearance needs to be extrapolated (e.g., if new images are acquired later in a longitudinal study after the previous time-points have been aligned). This will be investigated as part of future work.
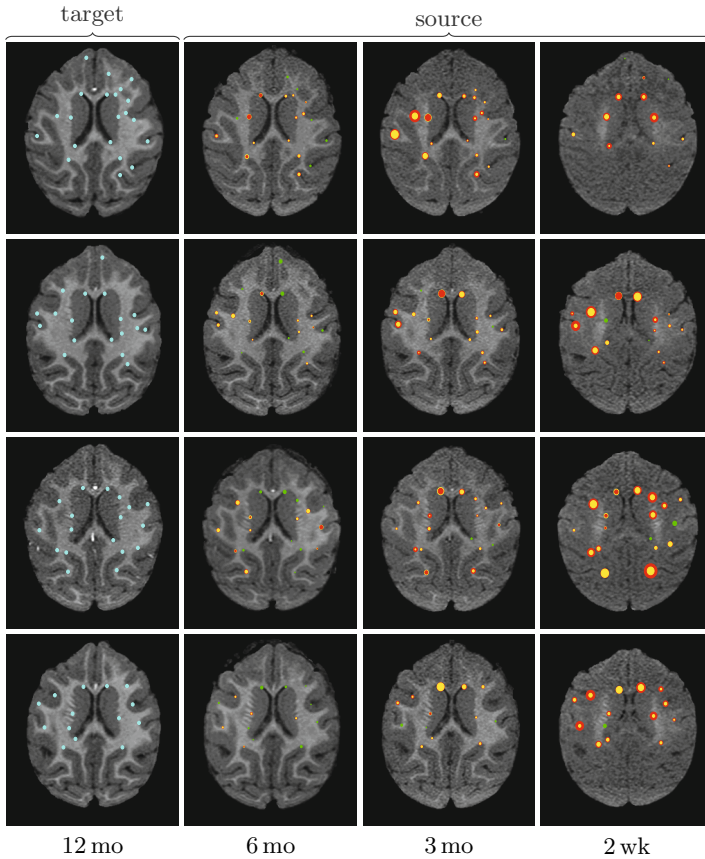
Fig. 5. Landmark registration error. Each row shows a single subject. The first column shows the target images and landmarks (cyan). Columns 2-4 are the source images and landmarks: each source landmark is marked as two circles (*red*: MI, *yellow*: mSM) with size proportional to registration error (smaller circle on top; *green*: both are equal). That is, the size of the circles is proportional to registration accuracy (smaller is more accurate) in that particular location.

Table 1. Landmark registration error (in voxels) between target $I_{12\mathrm{mo}}$ and source images $I_{2\mathrm{wk}}, I_{3\mathrm{mo}}, I_{6\mathrm{mo}}$ (significance level is $\alpha = 0.05$; significant results are highlighted).

	$I_{2\mathrm{wk}}$				$I_{3\mathrm{mo}}$				$I_{6\mathrm{mo}}$			
	mean	*std*	50th	90th	*mean*	*std*	50th	90th	*mean*	*std*	50th	90th
MI	1.76	1.09	1.60	3.22	1.07	0.77	0.84	2.04	0.66	0.46	0.56	1.22
mSM	**1.15**	**0.84**	**0.93**	**2.24**	**0.74**	**0.57**	**0.54**	**1.66**	0.61	0.39	0.50	1.18

5   Conclusions

We proposed a new model-based similarity measure which allows the deformable registration of longitudinal images with appearance change. This method can account for the intensity change over time and enables the registration method to recover the deformation due only to changes in morphology. We compared the model-based approach to mutual information and demonstrated that it can achieve higher accuracy than mutual information in cases when there is a large appearance change between source and target images. We used a logistic model of intensity change and an elastic deformation model, however, the formulation is general and can be used with any other appearance or deformation model. In the future we will investigate the use of prior models to inform the estimation step in regions with high uncertainty (e.g., due to poor initial alignment), in addition to the effect of various intensity models on the registration accuracy.

References

1. Barkovich, A.J., Kjos, B.O., Jackson, D.E., Norman, D.: Normal maturation of the neonatal and infant brain: MR imaging at 1.5T. Radiology 166, 173–180 (1988)
2. Broit, C.: Optimal registration of deformed images. Ph.D. thesis, University of Pennsylvania (1981)
3. Csapo, I., Davis, B., Shi, Y., Sanchez, M., Styner, M., Niethammer, M.: Temporally-Dependent Image Similarity Measure for Longitudinal Analysis. In: Dawant, B.M., Christensen, G.E., Fitzpatrick, J.M., Rueckert, D. (eds.) WBIR 2012. LNCS, vol. 7359, pp. 99–109. Springer, Heidelberg (2012)
4. Dobbing, J., Sands, J.: Quantitative growth and development of human brain. Archives of Disease in Childhood 48(10), 757–767 (1973)
5. Durrleman, S., Pennec, X., Trouvé, A., Gerig, G., Ayache, N.: Spatiotemporal Atlas Estimation for Developmental Delay Detection in Longitudinal Datasets. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009, Part I. LNCS, vol. 5761, pp. 297–304. Springer, Heidelberg (2009)
6. Fekedulegn, D., Siurtain, M.P.M., Colbert, J.J.: Parameter estimation of nonlinear growth models in forestry. Silva Fennica 33(4), 327–336 (1999)
7. Fishbaugh, J., Durrleman, S., Gerig, G.: Estimation of Smooth Growth Trajectories with Controlled Acceleration from Time Series Shape Data. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part II. LNCS, vol. 6892, pp. 401–408. Springer, Heidelberg (2011)
8. Friston, K., Ashburner, J., Frith, C., Poline, J., Heather, J.D., Frackowiak, R.: Spatial registration and normalization of images. Human Brain Mapping 2, 165–189 (1995)
9. Loeckx, D., Slagmolen, P., Maes, F., Vandermeulen, D., Suetens, P.: Nonrigid image registration using conditional mutual information. IEEE Transactions on Medical Imaging 29(1), 19–29 (2010)

10. Miller, M.I., Younes, L.: Group actions, homeomorphisms, and matching: A general framework. International Journal of Computer Vision 41(1/2), 61–84 (2001)
11. Roche, A., Guimond, A., Ayache, N., Meunier, J.: Multimodal Elastic Matching of Brain Images. In: Vernon, D. (ed.) ECCV 2000. LNCS, vol. 1843, pp. 511–527. Springer, Heidelberg (2000)
12. Sampaio, R.C., Truwit, C.L.: Myelination in the developing brain. In: Handbook of Developmental Cognitive Neuroscience, pp. 35–44. MIT Press (2001)
13. Studholme, C., Drapaca, C., Iordanova, B., Cardenas, V.: Deformation-based mapping of volume change from serial brain mri in the presence of local tissue contrast change. IEEE Transactions on Medical Imaging 25(5), 626–639 (2006)
14. Styner, M., Knickmeyer, R., Coe, C., Short, S.J., Gilmore, J.: Automatic regional analysis of dti properties in the developmental macaque brain. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, vol. 6914 (2008)

Cortical Folding Analysis
on Patients with Alzheimer's Disease
and Mild Cognitive Impairment

David M. Cash[1,2], Andrew Melbourne[1], Marc Modat[1], M. Jorge Cardoso[1],
Matthew J. Clarkson[1], Nick C. Fox[2], and Sebastien Ourselin[1,2]

[1] Centre for Medical Image Computing, University College of London, UCL*
[2] Dementia Research Centre, University College of London, UCL

Abstract. Cortical thinning is a widely used and powerful biomarker
for measuring disease progression in Alzheimer's disease (AD). However,
there has been little work on the effect of atrophy on the cortical folding
patterns. In this study, we examined whether the cortical folding could
be used as a biomarker of AD. Cortical folding metrics were computed
on 678 patients from the Alzheimer's Disease Neuroimaging Initiative
(ADNI) cohort. For each subject, the boundary between grey matter and
white matter was extracted using a level set technique. At each point on
the boundary two metrics characterising folding, curvedness and shape
index, were generated. Joint histograms using these metrics were calcu-
lated for five regions of interest (ROIs): frontal, temporal, occipital, and
parietal lobes as well as the cingulum. Pixelwise statistical maps were
generated from the joint histograms using permutations tests. In each
ROI, a significant reduction was observed between controls and AD in
areas associated with the sulcal folds, suggesting a sulcal opening asso-
ciated with neurodegeneration. When comparing to MCI patients, the
regions of significance were smaller but overlapping with those regions
found comparing controls to AD. It indicates that the differences in cor-
tical folding are progressive and can be detected before formal diagnosis
of AD. Our preliminary analysis showed a viable signal in the cortical
folding patterns for Alzheimer's disease that should be explored further.

1 Introduction

Alzheimer's disease (AD) is a progressive neurodegenerative disorder that affects
nearly 1 in 8 people over 65 in the U.S.[1], and the prevalence is expected to
increase in the near future. The hallmark pathology of AD are amyloid plaques
and neurofibrillary tangles, which result in neuronal dysfunction. This ultimately
leads to cell death which can be observed macroscopically in structural MRI as
brain atrophy. Numerous imaging biomarkers have been proposed for the robust

and accurate measurement of this atrophy process [2,3]. One of the most popular methods has been to measure cortical thickness [4,5]. In this method, the cortical ribbon is extracted and correspondences are established between the grey matter (GM) to white matter (WM) boundary and the pial surface, from which thickness measurements can then be computed. Cortical thickness is a biomarker that has been shown to separate patients with AD from healthy controls [5], and to find a pattern of thinning [6] that closely resembles what is observed through post-mortem studies [7]. In [8], there were some regions, primarily in the temporal lobe that appeared to thin at greater rates if the subject showed signs of Alzheimer's disease pathology according to cerebrospinal fluid markers of amyloid beta and tau.

Besides cortical thickness, the cortical folding pattern has been analysed for numerous applications. Manual histology based methods frequently use a gyrification index (GI) defined on slices measuring the ratio of the length of the grey matter to that of the shortest route around the coronal surface [9,10]. Computational methods using structural MRI require an accurate underlying segmentation and are in principle defined in 3D. [11] made use of a level set routine in order to assess the scaling relation of cortical volume to cortical surface area and developed summary statistics of the cortical surface by collecting complementary values of curvature into two-dimensional histograms. The representation of the surface is such that their histograms have two visible peaks distinguishing gyri from sulci. The histograms can be assessed using pixelwise statistical tests. In [12] the authors analysed the cortical surface by defining a mesh on the surface of the cortex and defining a mean curvature found from nodes within a local region. This study of healthy adults found strong gender dimorphism although the impact of overall subject size was unclear. The longitudinal cortical folding pattern of neonates has also been investigated by [13] in which after a semi-automatic segmentation, the cortical surface was delineated by a surface mesh. The authors analysed the appearance and formation of major sulci over 26-36 weeks gestational age making use of a sulcation ratio of the area of major sulci in relation to total cortical area.

Despite the volume of work on cortical folding, there has been very little focus on examining how it is affected by the process of neurodegeneration. King *et al.* [14] used fractal dimension analysis and gyrification index in addition to cortical thickness measurements to study the cortical ribbon in patients with AD. In this study, we applied the curvature metrics defined in [11] to determine if there are group wise differences in cortical folding patterns between normal controls (NC), mild cognitively impaired (MCI), and patients with AD.

2   Methods

2.1   Data

The data for this study consisted of a total of 678 participants, divided into 4 groups: controls (n=199, age 76.0 (5.1), 47% female, MMSE 29.1 (1.0)), MCI patients stable after followup of up to 36 months (n=192, age 75.2 (7.1), 34%

female, MMSE 27.0 (1.8)), MCI patients who converted to AD during followup (n=140, age 74.3 (7.4), 39% female, MMSE 27.1 (1.7)) and 147 subjects diagnosed with AD (mean age 75.3 (7.3), 47% female, MMSE 23.4 (1.9)). All subjects were selected from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database [15][1]. A 1.5T scan with sufficient image quality from each subject's baseline visit was downloaded from the ADNI database. All images went through the standard pre-processing pipeline of bias correction, correction of B_1 inhomogeneity, and correction of geometric distortion [16]. The whole brain was manually delineated through a semi-automated process [17], which was used for further calculations.

2.2   Tissue Classification and Brain Parcellation

The pre-processed images were masked using the brain delineations and then segmented using an expectation-maximisation based algorithm [18] into five tissue classes: cortical GM, WM, external CSF, deep GM and internal CSF. The image was also parcellated into anatomical regions of interest. A library of 30 atlases manually labelled with 83 anatomical regions [19] was used as a basis for the parcellations. Each of the atlases was propagated via a nonlinear registration to provide a set of candidate segmentations for the new image. These candidate parcellations were combined using a label fusion strategy based on the STAPLE method [20]. It was adapted to use only a subset of the most similar candidate segmentations chosen for each voxel by a local image similarity metric [21].

2.3   Boundary Surface Generation

The segmentation method allows an inspection of the shape of the boundary between grey and white matter surfaces similar to [11]. In our work we optimise the boundary between GM and WM using an evolving level set, solving a partial differential equation (1) for the function $\phi(\mathbf{x}, t)$, consisting of a weighted signed distance function from the boundary, d, and a curvature function, c, that serves as a regularisation term. The level set is initialised at $p(\text{WM}) = 0.5$, that provides a signed distance function from the GM/WM boundary.

$$\frac{\delta\phi}{\delta t} + (\lambda_d d(\mathbf{x}, t) - \lambda_c c(\mathbf{x}, t)) ||\nabla\phi(\mathbf{x}, t)|| = 0 \qquad (1)$$

The outer boundary of GM can be defined by evolving the initial GM/WM surface outward to the $p(\text{WM}) + p(\text{GM}) = 0.5$ boundary. The ratio of the inner surface (GM/WM interface) area to the outer surface area (GM/CSF interface) can be used as an approximation of the gyrification index [9].

2.4   Curvature Measurements

We define measures of curvature on this implicit surface, similar to the method used in [11]. The local Hessian matrix of second order derivatives can be found

---

[1] www.adni-info.org

at each point of this implicit surface. Hence, at every point on the surface we can use the eigenvalues of the local Hessian matrix, κ_1 and κ_2 (with $\kappa_1 > \kappa_2$) to summarise the local shape. Explicitly we define the shape index, S (2), describing how cup-like or saddle-like the surface is,

$$S = \frac{2}{\pi}\tan^{-1}\left(\frac{\kappa_2 + \kappa_1}{\kappa_2 - \kappa_1}\right) \qquad (2)$$

and a curvature, C, describing the distortion of the surface relative to that of a flat sheet ([22]): $C = \sqrt{(\kappa_1)^2 + (\kappa_2)^2}$. Intuitively C and S are complementary in the sense that they are a polar representation of the eigenvalue sum $(\kappa_1 + \kappa_2)$ and difference $(\kappa_2 - \kappa_1)$. The value of C is corrected for brain volume using the strategy described in [11], where each value of C is divided by a correction factor: $\beta_i^3 = \overline{ICV}/ICV_i$, where ICV is the intracranial volume.

The distribution of C and S can be summarised in a 2D histogram [11], in which empirically, the resulting histograms have two clusters corresponding to negative (sulci) and positive shape index (gyri). Joint histograms were constructed from these pairs of values at every point on the level set. We used a joint histogram of 32 bins along each axis, with a range for curvature from 0 to 1.5 and shape index from -1 to 1. In order to provide more spatial localisation to the measurements, the joint histograms were built over various regions of interest (ROI) as defined by the parcellation. Individual regions were combined into 5 meta regions: the cingulum, frontal, occipital, parietal, and temporal lobes. From the joint histogram, a bivariate probability density function (PDF) was estimated using an adaptive kernel density estimation technique [23] based on the discrete cosine transform.

2.5   Statistical Analysis

Pixelwise statistical tests were performed on the joint histogram for each ROI's from all of the subjects. The resulting t-statistic values were corrected for multiple comparisons using a non-parameteric method [24]. 5000 permutations were performed on 12 t-tests for each region. These t-tests represent forward (i.e. control > AD) and reverse (i.e. AD > control) comparisons between each of the four groups. Age and gender were accounted for in the design matrix as covariates.

3   Results

Mean GI values, adjusted for TIV, gender and age, for the four groups are shown in Fig. 1. There were significant differences for controls vs. MCI-Stable ($p = 0.027$), controls vs. MCI-Converters ($p = 0.011$), controls vs. AD ($p < 0.001$), and MCI Stable vs AD ($p = 0.039$). There was no difference between the two MCI subgroups ($p = 0.617$) or MCI Converters vs. AD ($p = 0.147$).

Figure 2 shows the average histogram for each ROI overlaid with colour coded areas when a t-test reached significance at ($p < 0.05$) after correcting for multiple comparisons. In all five ROIs, there are regions of statistical significance (red, top
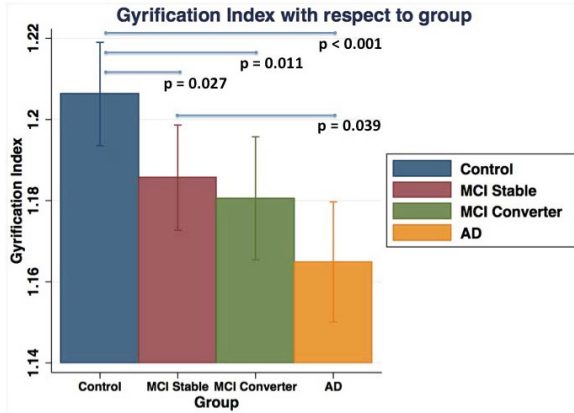
Fig. 1. Mean (95% confidence intervals) of gyrification index for each of the groups. The values have been adjusted for age, gender and total intracranial volume.

row) that indicate where the values in the histogram are higher in controls than AD. These occur mostly in areas near the peak of the joint histogram, where S is negative and C$\approx 0.2 - 0.5$ mm^{-1}. For the cingulum, there is also a region at a similar curvature but where S is positive. Comparisons involving MCI subgroups resulted in smaller regions primarily contained within the control $>$ AD region. Significant regions were detected for control $>$ MCI Converters (all ROIs except frontal), controls $>$ MCI Stable (frontal, temporal, parietal), MCI Stable $>$ AD (temporal, parietal), and MCI Converter $>$ AD (temporal).

In all five ROIs, there were areas of significance (red regions, bottom row) when using the reverse comparison, AD $>$ Control. For each ROI except the cingulum, these occurred in two areas of the histogram: (1) upper left, where S is negative and C > 0.65 mm^{-1} and (2) far right, where S is close to $+1$ and C ≈ 0.2 mm^{-1}. In the cingulum, there were also significant regions in the very lower left: where S is negative and C< 0.175 mm^{-1}. Similar smaller overlapping regions were again observed when performing comparisons with MCI subgroups including: MCI Converter $>$ Control (temporal, occipital), MCI Stable $>$ Control (cingulum), AD $>$ MCI Stable (frontal,temporal parietal).

4   Discussion

The cortical folding measurements in this paper track changes between disease groups that appear to be related to neurodegeneration. In Fig. 1, there are statistically significant differences between a summary statistic derived from the cortical folding. This value is higher in controls than MCI and AD subjects, which was also observed by King *et al.* [14], indicating the complexity of cortical folding is decreasing due to the disease process.

The main areas of significance in the forward comparison are near one of the histogram peaks where S is negative and C is moderate, typically attributed
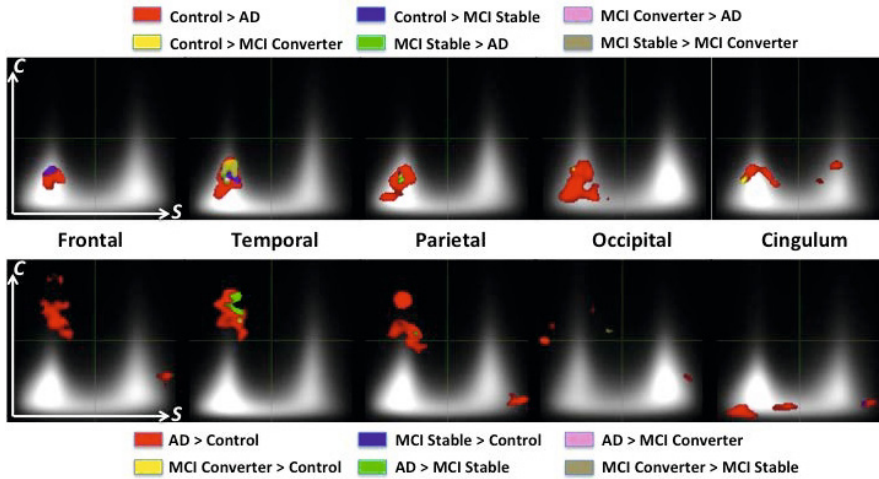
Fig. 2. Statistical maps for each region of interest. The X axis represents the shape index (range -1 to +1, unitless) and the Y axis is the curvature value (range 0 to 1.5mm^{-1}). The top row has the forward comparisons (*i.e.* control > AD) and the bottom row the reverse comparisons (*i.e.* AD > control). Areas of statistical significance ($p < 0.05$) are overlaid on the average joint histogram as colour coded in the legend.

to sulcal areas. As these regions are greater in healthier patients (i.e Control > MCI > AD), this suggests that the sulci are widening and thus the curvature is reducing, thus flattening the peak of the histogram. There are also regions of significance when performing reverse comparisons. These occur both in extremely high curvature areas with negative shape indices or extremely low curvature areas with positive shape indices (and also in negative shape indices in the case of the cingulum). The high curvature areas have much lower values (5–30%) in the PDF compared to the areas representing the sulcal peaks. It could either be that at this image resolution, the level set is unable to accurately acquire these areas of high curvature in healthy controls until there is further sulcal opening, or that neurodegeneration is further opening deep sulcal spaces and increasing cortical folding complexity. The low curvature areas being higher in affected patients seems to support the idea that neurodegeneration is reducing the curvature in the cortical folds. It is encouraging that the comparisons involving MCI subgroups result in smaller clusters that overlap with regions associated with controls vs. AD. This appears to suggest the measurement tracks with disease progression. The temporal lobe appears to show the greatest discriminant power, as it showed the most discrimination between the four groups. This is in agreement with many studies which indicate structures in the temporal lobe are some of the earliest to change.

It will be important to compare this method with cortical thickness measurements to determine if there is a correlation between the two measurements. As there is significant cortical variability between subjects, it would be difficult to apply this method on a vertex or voxel wide basis as is done in cortical thickness.

However, it would be worthwhile to compute the joint histograms over smaller ROIs defined, such as individual cortical regions from the parcellation. Often in clinical trials, a summary statistic is more desirable than statistical maps to asses the effectiveness of treatment. Therefore, it is imperative that effective summary statistics, similar to the gyrification index, are developed from these metrics. As the level set can be evolved outwards to the pial surface, it would be worthwhile to test the effect of computing the curvature metrics from different boundaries, such as the outer pial surface or a midline between the two boundaries, as King found different levels of correlation between their fractal dimension values and the cortical thickness depending on which boundary was used.

5 Conclusion

From this preliminary study, there is evidence that the cortical folding pattern could be a useful biomarker in disease progression of Alzheimer's disease. While primarily applied to areas of development (such as neonatal applications), these markers also apply to neurodegeneration as well. Further work to refine the method and determine longitudinal changes should be pursued.

References

1. Alzheimer's Association, Alzheimer's facts and figures
2. Holland, D., Dale, A.M.: Nonlinear registration of longitudinal images and measurement of change in regions of interest. Med. Image Anal. 15(4), 489–497 (2011)
3. Leung, K.K., Ridgway, G.R., Ourselin, S., Fox, N.C.: Consistent multi-time-point brain atrophy estimation from the boundary shift integral. NeuroImage 59(4), 3995–4005 (2011)
4. Fischl, B., Dale, A.M.: Measuring the thickness of the human cerebral cortex from magnetic resonance images. PNAS 97(20), 11050–11055 (2000)
5. Acosta, O., Bourgeat, P., Zuluaga, M., Fripp, J., Salvado, O., Ourselin, S.: Automated voxel-based 3D cortical thickness measurement in a combined Lagrangian-Eulerian PDE approach using partial volume maps. Med. Image Anal. 13(5), 730–743 (2009)
6. Dickerson, B.C., Bakkour, A., Salat, D.H., Feczko, E., Pacheco, J., Greve, D.N., Grodstein, F., Wright, C.I., Blacker, D., Rosas, H.D., Sperling, R.A., Atri, A., Growdon, J.H., Hyman, B.T., Morris, J.C., Fischl, B., Buckner, R.L.: The cortical signature of Alzheimer's disease: regionally specific cortical thinning relates to symptom severity in very mild to mild AD dementia and is detectable in asymptomatic amyloid-positive individuals. Cerebral Cortex 19(3), 497–510 (2009)
7. Braak, H., Braak, E.: Neuropathological staging of Alzheimer-related changes. Acta Neuropathologica 82(4), 239–259 (1991)
8. Desikan, R.S., Sabuncu, M.R., Schmansky, N.J., Reuter, M., Cabral, H.J., Hess, C.P., Weiner, M.W., Biffi, A., Anderson, C.D., Rosand, J., Salat, D.H., Kemper, T.L., Dale, A.M., Sperling, R.A., Fischl, B.: Selective disruption of the cerebral neocortex in Alzheimer's disease. PloS One 5(9), e12853 (2010)
9. Zilles, K., Armstrong, E., Schleicher, A., Kretschmann, H.J.: The human pattern of gyrification in the cerebral cortex. Anat. Embryol. 179, 173–179 (1988)

10. Armstrong, E., Schleicher, A., Omran, H., Curtis, M., Zilles, K.: The ontogeny of human gyrification. Cerebral Cortex 5(1), 56–63 (1995)
11. Awate, S.P., Yushkevich, P.A., Song, Z., Licht, D.J., Gee, J.C.: Cerebral cortical folding analysis with multivariate modeling and testing: Studies on gender differences and neonatal development. NeuroImage 53(2), 450–459 (2010)
12. Luders, E., Thompson, P.M., Narr, K.L., Toga, A.W., Jancke, L., Gaser, C.: A curvature-based approach to estimate local gyrification on the cortical surface. NeuroImage 29(4), 1224–1230 (2006)
13. Dubois, J., Benders, M., Borradori-Tolsa, C., Cachia, A., Lazeyras, F., Ha-Vinh Leuchter, R., Sizonenko, S.V., Warfield, S.K., Mangin, J.F., Hüppi, P.S.: Primary cortical folding in the human newborn: an early marker of later functional development.. Brain 131(8), 2028–2041 (2008)
14. King, R.D., Brown, B., Hwang, M., Jeon, T., George, A.T.: Fractal dimension analysis of the cortical ribbon in mild Alzheimer's disease. NeuroImage 53(2), 471–479 (2010)
15. Mueller, S.G., Weiner, M.W., 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 Clinics of North America 15(4), 869–877 (2005)
16. Jack, C.R., Bernstein, M.A., Fox, N.C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P.J., Whitwell, J.L., Ward, C., Dale, A.M., Felmlee, J.P., Gunter, J.L., Hill, D.L.G., Killiany, R., Schuff, N., Fox-Bosetti, S., Lin, C., Studholme, C., DeCarli, C.S., Krueger, G., Ward, H.A., Metzger, G.J., Scott, K.T., Mallozzi, R., Blezek, D., Levy, J., Debbins, J.P., Fleisher, A.S., Albert, M., Green, R., Bartzokis, G., Glover, G., Mugler, J., Weiner, M.W.: The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. JMRI 27(4), 685–691 (2008)
17. Freeborough, P.A., Fox, N.C., Kitney, R.I.: Interactive algorithms for the segmentation and quantitation of 3-D MRI brain scans. Computer Methods and Programs in Biomedicine 53(1), 15–25 (1997)
18. Cardoso, M.J., Clarkson, M.J., Ridgway, G.R., Modat, M., Fox, N.C., Ourselin, S.: LoAd: a locally adaptive cortical segmentation algorithm. NeuroImage 56(3), 1386–1397 (2011)
19. Hammers, A., Chen, C.H., Lemieux, L., Allom, R., Vossos, S., Free, S.L., Myers, R., Brooks, D.J., Duncan, J.S., Koepp, M.J.: Statistical neuroanatomy of the human inferior frontal gyrus and probabilistic atlas in a standard stereotaxic space. Human Brain Mapping 28(1), 34–48 (2007)
20. Warfield, S., Zou, K., Wells, W.: Simultaneous truth and performance level estimation (staple): an algorithm for the validation of image segmentation. IEEE Trans. Med. Im. 23(7), 903–921 (2004)
21. Cardoso, M.J., Modat, M., Ourselin, S., Keihaninejad, S., Cash, D.: Multi-STEPS: Multi-label similarity and truth estimation for propagated segmentations. In: 2012 IEEE Workshop on Mathematical Methods in Biomedical Image Analysis, pp. 153–158. IEEE (2012)
22. Koenderink, J.J., van Doorn, A.J.: Surface shape and curvature scales. Image and Vision Computing 10(8), 557–564 (1992)
23. Botev, Z.I., Grotowski, J.F., Kroese, D.P.: Kernel density estimation via diffusion. The Annals of Statistics 38(5), 2916–2957 (2010)
24. Nichols, T.E., Holmes, A.P.: Nonparametric permutation tests for functional neuroimaging: a primer with examples. Human Brain Mapping 15(1), 1–25 (2002)

Inferring Group-Wise Consistent Multimodal Brain Networks via Multi-view Spectral Clustering

Hanbo Chen[1], Kaiming Li[1,2], Dajiang Zhu[1], Tuo Zhang[1,2], Changfeng Jin[3],
Lei Guo[2], Lingjiang Li[3], and Tianming Liu[1]

[1] Department of Computer Science and Bioimaging Research Center,
The University of Georgia, Athens, GA, USA
[2] School of Automation, Northwestern Polytechnical University, Xi'an, China
[3] Department of Psychiatry, The Mental Health Institute,
The Second Xiangya Hospital, Central South University, Changsha, China

Abstract. Quantitative modeling and analysis of structural and functional brain networks based on diffusion tensor imaging (DTI)/functional MRI (fMRI) data has received extensive interest recently. However, the regularity of these structural or functional brain networks across multiple neuroimaging modalities and across individuals is largely unknown. This paper presents a novel approach to infer group-wise consistent brain sub-networks from multimodal DTI/fMRI datasets via multi-view spectral clustering of cortical networks, which were constructed on our recently developed and extensively validated large-scale cortical landmarks. We applied the proposed algorithm on 80 multimodal structural and functional brain networks of 40 healthy subjects, and obtained consistent multimodal brain sub-networks within the group. Our experiments demonstrated that the derived brain sub-networks have improved inter-modality and inter-subject consistency.

Keywords: DTI, fMRI, multimodal brain connectivity, multi-view clustering.

1 Introduction

Studying structural/functional brain networks via DTI/fMRI has attracted increasing interest recently due to its potential in elucidating fundamental architectures and principles of the brain [1]. E.g., in [2], Beckmann applied PICA to analysis fMRI; in [3], the functional connectivity in resting brain is studied to infer default mode network. In many previous studies, structural and functional brain networks are typically examined separately, leaving their relationship largely unknown. In addition, the regularity of the structural or functional brain networks across multiple neuroimaging modalities and across different brains has rarely been investigated. Essentially, better quantitative characterization of the relationship between multimodal brain networks and its consistency across individuals could significantly advance our understanding of the human brain architectures [4].

In response to this issue, this paper presents a novel approach to infer group-wise consistent brain networks from multimodal DTI/fMRI datasets via multi-view spectral clustering of large-scale cortical landmarks and their connectivities. Based on our

recently developed and validated large-scale connectivity-based cortical landmarks [5] as network nodes, we constructed both structural and functional brain networks from multimodal DTI/fMRI data of 40 healthy brains. Then, we applied an effective multi-view spectral clustering algorithm [6] on these 80 multimodal structural and functional brain networks to derive consistent multimodal brain sub-networks.

The prominent advantage of multi-view spectral clustering methodology is that it can effectively deal with heterogeneous features by maximization of the mutual agreement across multimodal clusters in different views [6]. In this work, we considered each structural or functional network in a subject as a separate view of the studied large-scale network. We modeled the clustering of group-wise consistent multimodal brain sub-networks in a unified multi-view clustering framework, by which the substantial variability of large-scale brain networks across modalities (DTI and R-fMRI) and different individuals (40 subjects) is modeled and handled by the powerful multi-view spectral clustering method. This is the major methodological novelty and contribution of this paper.

Our experimental results have shown that the derived brain sub-networks via the multi-view spectral clustering method have improved inter-modality predictability and consistency in comparison with clustering results by single modality, and have improved inter-subject consistency. More importantly, they are consistent with current neuroscience knowledge and better explain the relationship between brain structure and function. Our study provides novel insights on inferring reliable and reproducible multimodal networks and network-based signatures for the elucidation of brain function and dysfunction in the future.

2   Methods

Our computational pipeline is summarized in Fig. 1. Both structural and functional brain networks (Fig. 1a-b) were constructed from DTI and resting state fMRI (R-fMIR) data of the same group of subjects based on our recently validated 358 cortical landmarks [5]. The joint connectivity matrix (Fig. 1c) is then computed via multi-view spectral clustering algorithm. Then, the clustering procedure generates group-wise consistent multimodal sub-networks (Fig. 1d), which are then projected back to the original 358 cortical landmark space for visualization and validation (Fig. 1e). Finally, we quantitatively compared different schemes in clustering brain network.
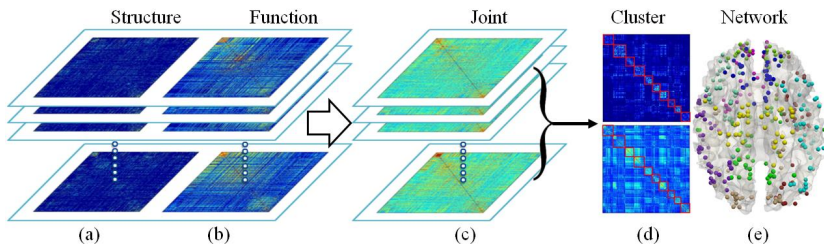
Fig. 1. The computational pipeline of proposed approach. (a) Structural connectivity matrix. (b) Functional connectivity matrix. (c) Joint connectivity matrix obtained via multi-view clustering algorithm. (d) Group-wise clustering. (e) Cluster sub-network.

2.1 Multimodal Brain Network Construction

A prerequisite to perform multi-view clustering of structural and functional networks is that the network nodes should possess correspondences across different modalities and individual brains. Recently we created and validated 358 cortical landmarks that have intrinsically established structural and functional correspondences in different brains [5], providing natural and ideal nodes for brain network construction. In brief, each of the 358 cortical landmarks was optimized to possess group-wise consistent white matter fiber connection patterns, which have been demonstrated to be predictive of functional localizations in the brain [5, 7]. The neuroscience basis is that each brain's cytoarchitectonic region has a unique set of intrinsic axonal inputs and outputs, called the "connectional fingerprint" [8], which largely determines the functions that brain area can perform. In particular, the functional correspondences of these 358 cortical landmarks were extensively validated by functional brain networks derived from both task-based fMRI data and R-fMRI data [5].
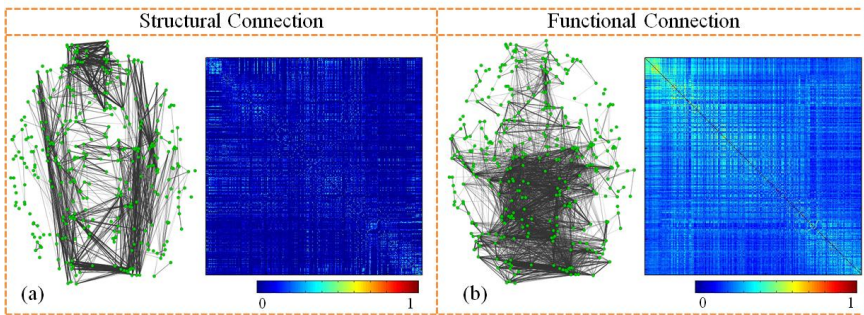
Fig. 2. An example of the constructed structural (a) and functional (b) networks. Both networks used the same set of 358 cortical ROIs as nodes. Each sub-figure shows a joint view of ROIs (green dot) and their connections (gray line) with corresponding connectivity matrix on right.

Based on these 358 cortical landmarks/ROIs (Fig. 2), we constructed both structural (Fig. 2a) and functional (Fig. 2b) networks for 40 healthy brains with multimodal DTI/R-fMRI data. We adopted R-fMRI data to construct the connectivity matrix of functional networks as follows. First, we performed brain tissue segmentation directly on DTI data [9], and used the gray matter segmentation map as a constraint for R-fMRI BOLD signal extraction. A principal component analysis was then conducted for the R-fMRI time series of all gray matter voxels within a ROI, and the first principal component was adopted as its representative R-fMRI BOLD signal. Then, the functional connection strength between ROIs is defined as the Pearson correlation of their R-fMRI BOLD signals.For the structural connectivity matrix, it was constructed from DTI data. Briefly, for each pair of ROIs, their connection strength is defined as the average FA (fractional anisotropy) value along the fiber bundle connecting the two ROIs. If there is no connecting fiber bundle between two ROIs, the connection strength is set to 0. An example of the constructed structural and functional networks is shown in Fig. 2.

2.2   Spectral Clustering

We used the spectral clustering to cluster the brain networks based on the connectivity matrix described above. In brief, spectral clustering is a technique that takes the advantage of the property of Laplacian of graph to reduce feature dimension for clustering. Specifically, k-means algorithm is used to cluster nodes on the first k eigenvectors of graph Laplacian with the smallest eigenvalue. The outline of the algorithm applied on our data is listed below. More details of the algorithm can be found in [10].

Input: Connectivity matrix S with size $n \times n$, where S_{ij} is the connection strength between nodes i and j. Number k of clusters to construct.

Output: Clusters of nodes.

1. Compute the normalized Laplacian $L = I - D^{-1/2}SD^{-1/2}$, where $D_{ii} = \sum_j S_{ij}$ is a diagonal matrix with graph degree on diagonal and 0 for the rest.
2. Compute the first k eigenvectors $u_1,...,u_k$ of L with smallest eigenvalue.
3. Let U denotes a $n \times k$ matrix containing $u_1,...,u_k$ that U_{ij} is the ith value of u_j. Normalize each row of U to obtain V.
4. i^{th} row in V corresponds to the ith node in S. Cluster nodes using each row of V as the feature vector via the k-means algorithm.

2.3   Co-training Approach for Multi-view Clustering

In our research problem, we have both structural connectivity and functional connectivity for large-scale brain network clustering. To find a common brain sub-networks across different modalities, the most intuitive way is to assign a weight to each view or modality. However, it is difficult to define optimal weights, especially when ROIs are unlabeled and there exists significant variability across modalities as shown in Figs. 2a and 2b. Thus, how to fuse these multimodal networks to achieve relatively consistent sub-networks becomes an important issue. Recently, a clustering method dubbed *multi-view clustering* has been developed to solve this type of problem [6, 11]. In this paper, we adopted a co-training approach based on spectral clustering [6] to maximize the agreement between structural network and functional network to find the consistent multimodal sub-networks of the human brain.

In spectral clustering, the reason why the first k eigenvectors can be used for clustering is that they contain the most discriminative information that can differentiate each cluster. By projecting the connectivity matrix to the space of the first k eigenvectors, the inner cluster details will be discarded and only essential information required for clustering retains. Thus, we can project the functional connectivity matrix to the space of the first k eigenvector of the Laplacian of structural connectivity matrix and then project it back and vice versa for structural connectivity matrix. By doing this iteratively, we can keep the common network between different modalities and discard the inconsistent information. Then, we combined feature vectors of both views and computed the Gaussian similarity between each ROI to obtain the joint connectivity matrix (Fig. 1c). The detailed algorithm is as follows [6].

Input: Connectivity matrix of two views S_1^0, S_2^0, number k of clusters to construct.

Output: Joint connectivity matrix S^*.

1. Compute the initial normalized Laplacian L_1^0, L_2^0 of each connectivity matrix, and the first k eigenvectors U_1^0, U_2^0 with smallest eigenvalue of L_1^0, L_2^0.
2. **for** $i = 1$ to *iter*
3. $\quad S_1^i = [U_2^{i-1}U_2^{i-1^T}S_1^{i-1} + (U_2^{i-1}U_2^{i-1^T}S_1^{i-1})^T]/2$
4. $\quad S_2^i = [U_1^{i-1}U_1^{i-1^T}S_2^{i-1} + (U_1^{i-1}U_1^{i-1^T}S_2^{i-1})^T]/2$
5. $\quad$ Compute Laplacian and corresponding first k eigenvectors U_1^i, U_2^i of S_1^i, S_2^i.
6. Normalize each row of U_1^i, U_2^i. Then combined normalized matrix to form a $n \times 2k$ matrix V.
7. i^{th} row in V correspond to the ith node in S_1, S_2.
8. Compute the Gaussian similarity $e^{\frac{-\|V(i)-V(j)\|_2^2}{\sigma}}$ between each row to obtain the joint connectivity matrix S^*.

3   Experimental Results

3.1   Data Acquisition, Preprocessing and Experiment Setup

Our experiment was performed on 40 healthy adults. Both DTI and R-fMRI were acquired for each subject. The parameters are as follows: R-fMRI: 64×64 matrix, 4 mm slice thickness, 220 mm FOV, 30 slices, TR = 2s; DTI: 256×256 matrix, 3 mm slice thickness, 240 mm FOV, 50 slices, 15 DWI volumes. Preprocessing including tissue segmentation, surface reconstruction and fiber tracking was performed with the same method in [5]. Then a set of large-scale, group-wise consistent ROIs were obtained on the cerebral cortex of each subject using method in [5]. The structural and functional connectivity matrices are then computed using method described in section 2.1. Examples of ROIs and connectivity matrices are shown in Fig. 2.

For comparison, the clustering has also been performed on structure connectivity matrix, function connectivity matrix and the mean of these two modalities. When only the structural connectivity matrix is used, the clustering only considers structural networks and thus the clustering of functional networks will be determined by the clustering result of the structural networks. Similarly, when the functional connectivity matrix is used, the clustering results of functional network will determine the grouping of structural networks. In the third scenario of using averaged structural and functional networks, we assigned equal weight to each view to obtain an average connectivity matrix. Then, the result sub-network will be based on both networks.

3.2   Clustering Results

The clustering results when the cluster number *k* is 10 are shown in Fig. 3. Each cluster is highlighted by a red box. By visualization, it is evident that the proposed multi-view clustering method generates more consistent sub-networks across modalities. When only structure network is used for clustering, the ROIs that have strong fiber connections are clustered together. However, the functional networks obtained have weaker connections between ROIs within each sub-network (Fig. 3c). Similarly, the clusters obtained by clustering functional network only tend to have weaker structural connections within each clustered sub-network (Fig. 3d). In contrast, the clusters by multi-view clustering and mean connectivity matrix considered both aspects (Fig. 3a-b). Importantly, our further quantitative analysis in section 3.3 will demonstrate that the clusters by multi-view clustering are more mutually consistent and predictable across different modalities and subjects.
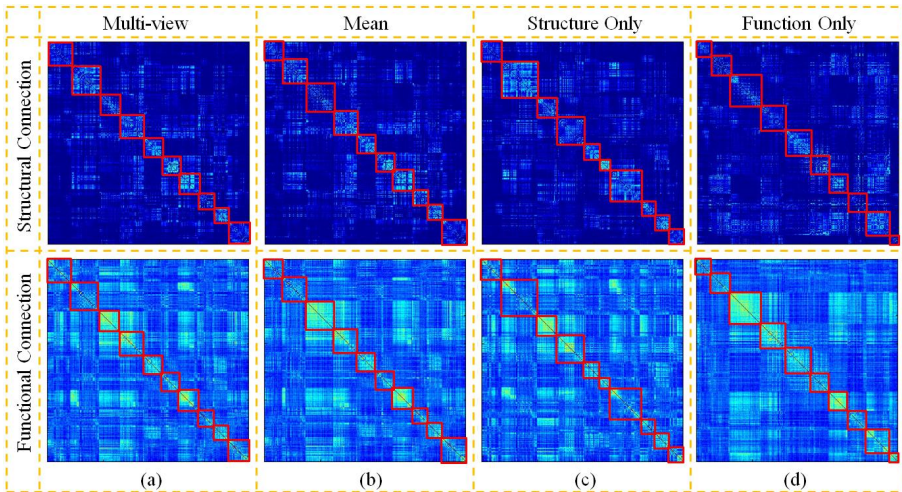
Fig. 3. Clustering results when *k* is 10. Each cluster of sub-network is highlighted by a red box. For each subfigure, the ID of the cluster from left to right is 1 to 10 respectively. (a)-(d) are results of clustering using multi-view, mean, structure only, and function only accordingly.

The clustered sub-networks by four schemes when k is ten are visualized in Fig. 4. Overall, there are certain similarities in the global patterns of clustered sub-networks by these four schemes (e.g. sub-network #3, #4, and #5 in Fig. 4). However, detailed examination of these sub-networks suggests that the multi-view clustering result has better agreement across modalities. For instance, according to the functional meta-analysis [5], both of the two ROIs highlighted by red box in Fig. 4 are involved in auditory perception and cognitive attention networks and should be clustered into the same network. But the clustering results based on structural connectivity matrix (Fig. 4c) and mean connectivity matrix (Fig. 4b) both failed in clustering these two ROIs into the same network.
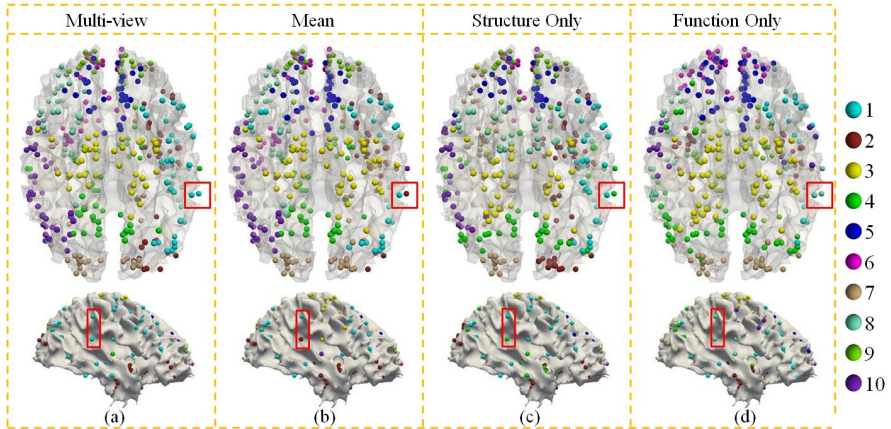
Fig. 4. Visualization of clustered brain sub-networks when k is 10. Each clustered sub-network is color-coded by a different color. The color scheme is on the right. The ID of each sub-network is the same with Fig. 3. (a)-(d) are results of clustering using multi-view, mean, structure only, and function only accordingly.

3.3   Quantitative Comparisons

In order to examine the clustering methods' sensitivity to cluster numbers and quantitatively compare these four schemes, we varied the cluster number k from 8 to 15 and measured the consistency between structural and functional sub-networks within each cluster. We used a linear regression model to predict functional network by structural network using the methods similar to that in [12]. Then, the regression residual is considered as the metric to assess how structural network is predictive and compatible with the functional network [12]. As shown in Fig. 5a, our results demonstrate that the regression residual by the proposed multi-view spectral clustering method is substantially smaller than other approaches, which indicates that the simultaneously clustered structural and functional brain sub-networks have improved predictability and consistency across modalities.
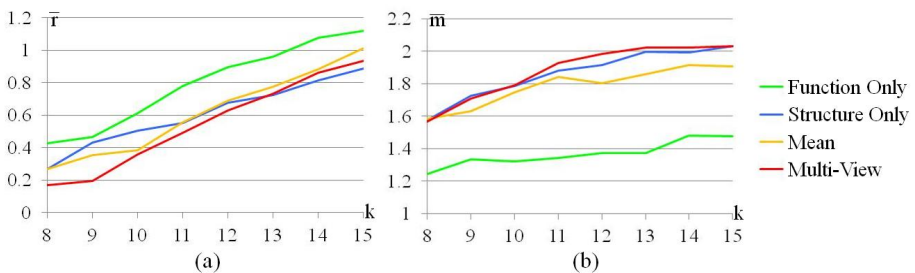
Fig. 5. Measurement of consistency between structural and functional networks of each clustered sub-network obtained by four different schemes. The horizontal axis is the number of clusters k, and the vertical axis is the consistency measurement. (a) Average residuals after linear regression from structural network to functional network within each cluster. (b) Average mutual information between each structural network and functional network within each cluster.

Furthermore, the mutual information between the structural and functional connectivity matrix within each clustered sub-network is computed and averaged across different brains. As mutual information is a quantitative measurement of the mutual dependence of two matrices, the relatively higher average mutual information within sub-networks by the proposed multi-view clustering result shown in Fig. 5b further indicates that this approach can generate more consistent multimodal brain networks across individuals compared with other three schemes.

4   Discussion and Conclusion

This paper presents a novel framework of clustering group-wise consistent multimodal brain networks based on DTI/R-fMRI datasets. The major methodological contribution of this work is modeling the inter-subject and inter-modality variations of brain networks by the multi-view spectral clustering algorithm. Experimental results demonstrated that the proposed multi-view clustering approach performs better than other schemes, and offered novel insights into the regularity of brain networks across modalities and individual brains. The effort in this paper can help construct both structurally and functionally meaningful and consistent brain networks in the future, which would have significant implications in basic and clinical neurosciences.

References

1. Bullmore, E., Sporns, O.: Complex brain networks: graph theoretical analysis of structural and functional systems. Nature Reviews Neuroscience 10, 186–198 (2009)
2. Beckmann, C.F., Smith, S.M.: Probabilistic independent component analysis for functional magnetic resonance imaging. IEEE Trans. Med. Imaging 23(2), 137–152 (2004)
3. Greicius, M.D., Krasnow, B., Reiss, A.L., Menon, V.: Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. Proc. Natl. Acad. Sci. U S A 100(1), 253–258 (2003)
4. Honey, C.J., Sporns, O., Cammoun, L., Gigandet, X., Thiran, J.P., Meuli, R., Hagmann, P.: Predicting human resting-state functional connectivity from structural connectivity. Proc. Natl. Acad. Sci. U S A 106(6), 2035–2040 (2009)
5. Zhu, D., Li, K., Guo, L., Jiang, X., Zhang, T., Zhang, D., Chen, H., Deng, F., Faraco, C., Jin, C., Wee, C.Y., Yuan, Y., Lv, P., Yin, Y., Hu, X., Duan, L., Hu, X., Han, J., Wang, L., Shen, D., Miller, L.S., Li, L., Liu, T.: DICCCOL: Dense Individualized and Common Connectivity-based Cortical Landmarks. Cerebral Cortex (in press, 2012)
6. Kumar, A., Daumé III, H.: A Co-training Approach for Multi-view Spectral Clustering. In: ICML (2011)
7. Zhang, T., Guo, L., Li, K., Jing, C., Yin, Y., Zhu, D., Cui, G., Li, L., Liu, T.: Predicting functional cortical ROIs via DTI-derived fiber shape models. Cerebral Cortex 22(4), 854–864 (2012)
8. Passingham, R.E., Stephan, K.E., Kötter, R.: The anatomical basis of functional localization in the cortex. Nat. Rev. Neurosci. 3(8), 606–616 (2002)
9. Liu, T., Li, H., Wong, K., Tarokh, A., Guo, L., Wong, S.T.: Brain tissue segmentation based on DTI data. NeuroImage 38(1), 114–123 (2007)
10. von Luxburg, U.: A Tutorial on Spectral Clustering. Statistics and Computing (2007)
11. Chaudhuri, K., Kakade, S.M., Livescu, K., Sridharan, K.: Multi-view clustering via canonical correlation analysis. In: ICML (2009)
12. Deligianni, F., Robinson, E., Beckmann, C.F., Sharp, D., Edwards, A.D.: Inference of functional connectivity from direct and indirect structural brain connections. In: ISBI (2011)

Test-Retest Reliability of Graph Theory Measures of Structural Brain Connectivity

Emily L. Dennis[1], Neda Jahanshad[1], Arthur W. Toga[2], Katie L. McMahon[3],
Greig I. de Zubicaray[5], Nicholas G. Martin[4], Margaret J. Wright[4,5],
and Paul M. Thompson[1]

[1] Imaging Genetics Center, Laboratory of Neuro Imaging, UCLA, CA, USA
[2] Laboratory of Neuro Imaging, UCLA, CA, USA
[3] Center for Advanced Imaging, Univ. of Queensland, Brisbane, Australia
[4] Queensland Institute of Medical Research, Brisbane, Australia
[5] School of Psychology, University of Queensland, Brisbane, Australia

Abstract. The human connectome has recently become a popular research topic in neuroscience, and many new algorithms have been applied to analyze brain networks. In particular, network topology measures from graph theory have been adapted to analyze network efficiency and 'small-world' properties. While there has been a surge in the number of papers examining connectivity through graph theory, questions remain about its test-retest reliability (TRT). In particular, the reproducibility of structural connectivity measures has not been assessed. We examined the TRT of global connectivity measures generated from graph theory analyses of 17 young adults who underwent two high-angular resolution diffusion (HARDI) scans approximately 3 months apart. Of the measures assessed, modularity had the highest TRT, and it was stable across a range of sparsities (a thresholding parameter used to define which network edges are retained). These reliability measures underline the need to develop network descriptors that are robust to acquisition parameters.

1   Introduction

Graph theory is increasingly used to analyze brain connectivity networks. Graph theory, a branch of mathematics concerned with the description and analysis of graphs, describes the brain as a set of nodes (brain regions) and edges (connections). Information on either structural or functional connectivity may be expressed in connectivity matrices, from which various network properties may be derived, such as clustering, efficiency, or small-world organization. Several of these measures have been shown to change during childhood development [1], and to be heritable [2], associated with specific genetic variants [3,4] and be altered in various neuropsychiatric disorders [5]. To date, only one study has examined the test-retest reliability (TRT) of these measures for *structural* networks, finding high reliability [6], but they did not examine different network sparsities. Results in the TRT of these measures in *functional* networks have been inconsistent. Low reliability [7], remarkably high reliability [8], and moderate reliability [9,10] have all been found. To define which connections are present in a network, often a sparsity threshold is

applied, to retain only those connections whose edge strengths exceed a given threshold, or to eliminate "weaker" connections. Telesford et al. [8], found that the reliability did not depend on the network sparsity, while Wang et al. [7], and Braun et al. [10], found it depended heavily on network sparsity and other user-selected network parameters. Zalesky et al. [11] suggested small-worldness or scale-freeness measures can change drastically depending on the scale of the parcellation, but few studies have assessed their reproducibility. As so many papers have been published using network measures, their reproducibility deserves further analysis.

We set out to examine the test-retest reliability of graph theory analyses of brain structural connectivity by scanning 17 young adults twice, over a 3-month interval, using high-angular resolution diffusion imaging (HARDI) at 4-Tesla. Other ongoing studies have assessed how connectivity matrices dependent on the scanner field strength, spatial, angular, and q-space resolution [12]. Here we assessed the reliability of commonly used network measures over a wide range of network sparsities, as well as inherently more robust measures integrated over different sparsity ranges.

2   Methods

2.1   Subjects

Our analysis included young adults aged 20-30 scanned twice with both MRI and DTI at 4T. Our analysis included a subset of a much larger cohort who was asked to return for a second scan, to assess reproducibility. Of these, some subjects were filtered out due to artifacts in their raw data or errors in tractography, leaving us with 26 subjects. Of these, 2 were statistical outliers on at least one graph theory metric (>3 SD from group mean), 5 had a large difference in the number of fibers tracked in scan 1 and scan 2 (difference of more than 33% in number of fibers in each scan), and 2 had a much larger interval between scan 1 and scan 2. After these subjects were filtered out, we were left with 17 subjects. Subjects were 12 female, 5 male, 100% Caucasian, mean age: 23.6 years, SD 1.47.

2.2   Scan Acquisition

Whole-brain anatomical and high angular resolution diffusion images (HARDI) were collected with a 4T Bruker Medspec MRI scanner. T1-weighted anatomical images were acquired with an inversion recovery rapid gradient echo sequence. Acquisition parameters were: TI/TR/TE = 700/1500/3.35ms; flip angle = 8 degrees; slice thickness = 0.9mm, with a 256x256 acquisition matrix. Diffusion-weighted images (DWI) were also acquired using single-shot echo planar imaging with a twice-refocused spin echo sequence to reduce eddy-current induced distortions. Acquisition parameters were optimized to provide the best signal-to-noise ratio for estimating diffusion tensors [13]. Imaging parameters were: 23cm FOV, TR/TE 6090/91.7ms, with a 128x128 acquisition matrix. Each 3D volume consisted of 55 2-mm thick axial slices with no gap and 1.79x1.79 mm^2 in-plane resolution. 105 images were acquired per subject: 11 with no diffusion sensitization (i.e., T2-weighted b_0 images) and 94

diffusion-weighted (DW) images (b = 1159 s/mm^2) with gradient directions evenly distributed on the hemisphere. Scan time for the HARDI scan was 14.2 min. The average scan interval was 101 days, SD 18 days.

2.3   Cortical Extraction and HARDI Tractography

Non-brain regions were automatically removed from each T1-weighted MRI scan, and from a T2-weighted image from the DWI set, using the FSL tool "BET" (FMRIB Software Library, http://fsl.fmrib.ox.ac.uk/fsl/). A trained neuroanatomical expert manually edited the T1-weighted scans to refine the brain extraction. All T1-weighted images were linearly aligned using FSL (with 9 DOF) to a common space [14] with 1mm isotropic voxels and a 220×220×220 voxel matrix. Raw diffusion-weighted images were corrected for eddy current distortions using the FSL tool, "eddy_correct" (http://fsl.fmrib.ox.ac.uk/fsl/). For each subject, the 11 eddy-corrected images with no diffusion sensitization were averaged, linearly aligned and resampled to a downsampled version of their corresponding T1 image (110×110×110, 2×2×2mm). Averaged b_0 maps were elastically registered to the structural scan to compensate for EPI-induced susceptibility artifacts. 35 cortical labels per hemisphere, as listed in the Desikan-Killiany atlas [15], were automatically extracted from all aligned T1-weighted structural MRI scans using FreeSurfer (http://surfer.nmr.mgh.harvard.edu/). As a linear registration is performed by the software, the resulting T1-weighted images and cortical models were aligned to the original T1 input image space and down-sampled using nearest neighbor interpolation (to avoid intermixing of labels) to the space of the DWIs. To ensure tracts would intersect labeled cortical boundaries, labels were dilated with an isotropic box kernel of width 5 voxels.

The transformation matrix from the linear alignment of the mean b_0 image to the T1-weighted volume was applied to each of the 94 gradient directions to properly re-orient the orientation distribution functions (ODFs). At each HARDI voxel, ODFs were computed using the normalized and dimensionless ODF estimator, derived for q-ball imaging (QBI) in [16]. We performed a recently proposed method for HARDI tractography [17] on the linearly aligned sets of DWI volumes using these ODFs. Tractography was performed using the Hough transform method as in [18]. Elastic deformations obtained from the EPI distortion correction, mapping the average b_0 image to the T1-weighted image, were then applied to the tracts' 3D coordinates for accurate alignment of the anatomy. Each subject's dataset contained 5000-10,000 useable fibers (3D curves). For each subject, a full 70×70 connectivity matrix was created. Each element described the proportion of the total number of fibers in the brain that connected a pair of labels; diagonal elements of the matrix describe the total number of fibers passing through a certain cortical region of interest. As these values were calculated as a proportion - they were normalized to the total number of fibers traced for each individual participant, so that results would not be skewed by raw fiber count.

2.4   Graph Theory Analyses

On the 70x70 matrices generated above, we used the Brain Connectivity Toolbox ([18]; https://sites.google.com/a/brain-connectivity-toolbox.net/bct/Home) to compute five standard measures of global brain connectivity - characteristic path length (CPL),

mean clustering coefficient (MCC), global efficiency (EGLOB), small-worldness (SW), and modularity (MOD) [18]. CPL is a measure of the average path length in a network; path length is the minimum *number* of edges that must be traversed to get from one node to another; it does not depend on the physical lengths of the fibers, only their network topology. MCC is a measure of how many neighbors of a given node are also connected to each other, as a proportion of the total number of connections in the network. EGLOB is inversely related to CPL: networks with a small average CPL are generally more efficient than those with large average CPL. SW represents the balance between network differentiation and network integration, calculated as a ratio of local clustering and characteristic path length of a node relative to the same ratio in a randomized network. We created 10 simulated random networks. MOD is the degree to which a system can be subdivided into smaller networks [19]. Figure 1 visualizes these measures in an example network.
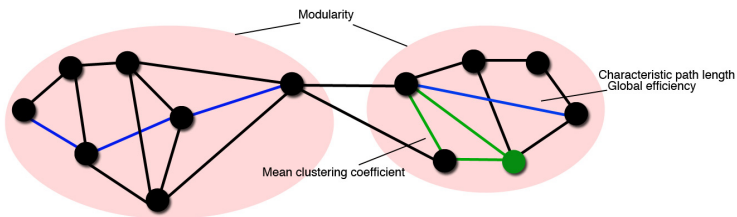
Fig. 1. Measures **of global connectivity**. Examples show network motifs that serve the basis of each measure. Adapted from the diagram in [18].

One step in binarized graph theory analyses is selecting a sparsity, which may be considered a thresholding operation on the edge strengths (here, fiber counts). The sparsity can alternatively be defined as the fraction of connections retained from the full network, so setting a sparsity level of 0.2 means that only the top 20% of connections (in this case, greatest numbers of fibers) are retained for calculations. The networks reconstructed at a given density will not be identical for any two people, but should be comparable as healthy people have highly similar white matter pathways, especially for the larger tracts. Selecting a single sparsity level may arbitrarily affect the network measures, so we typically compute measures at multiple sparsities, and integrate them across a range to generate more stable scores. We have previously used the range 0.2-0.3 to calculate and integrate these measures, as that range is biologically plausible [20] and more stable [4]. To determine whether the test-retest reliability varied across different sparsities we calculated these measures across the entire range (0-1 in 0.01 increments) as well as integrated across several smaller ranges (0.1-0.2, 0.2-0.3, 0.3-0.4, and 0.4-0.5, in 0.01 increments). We calculated these measures for the whole brain over these different sparsity ranges, and computed the area under the curve of those 11 data points to derive an integrated score for each measure.

2.5   Test-Retest Reliability Analyses

Test-retest reliability was measured by assessing the ICC (intraclass correlation coefficient) between graph theory measures generated from scan 1 matrices and scan 2 matrices. ICC is calculated according to the following formula:

$$ICC \ = \ \frac{MS_{Btwn} - MS_{Win}}{MS_{Btwn} + (k-1)MS_{Win}}$$

Here MS stands for mean squared deviation from the mean - within an individual or between individuals, and k stands for the number of scans for each subject (here $k=2$).

3   Results

3.1   Global Results

Test-retest reliability results for the full range of sparsities are shown for the 5 global measures in Figure 2. The very sparse measures (0-0.10) have low reliability, perhaps because different sets of nodes are retained between the first and second scans. Modularity (MOD) was most reliable network measure, with an r-value (reproducibility) between 0.35-0.65 for most of the range. Mean clustering coefficient (MCC) has an r-value mostly between 0.2-0.6, besides a sharp dip between sparsities 0.30-0.33. Small-worldness (SW) has an r-value that greatly fluctuates between 0.1-0.7, with many dips and peaks. Characteristic path length (CPL) and global efficiency (EGLOB) both are rather unreliable until around a sparsity of 0.25, at which point they both have r-values mostly between 0.3-0.6; the data depend heavily on the sparsity: note the sharp dip in reliability between sparsities 0.3-0.31.
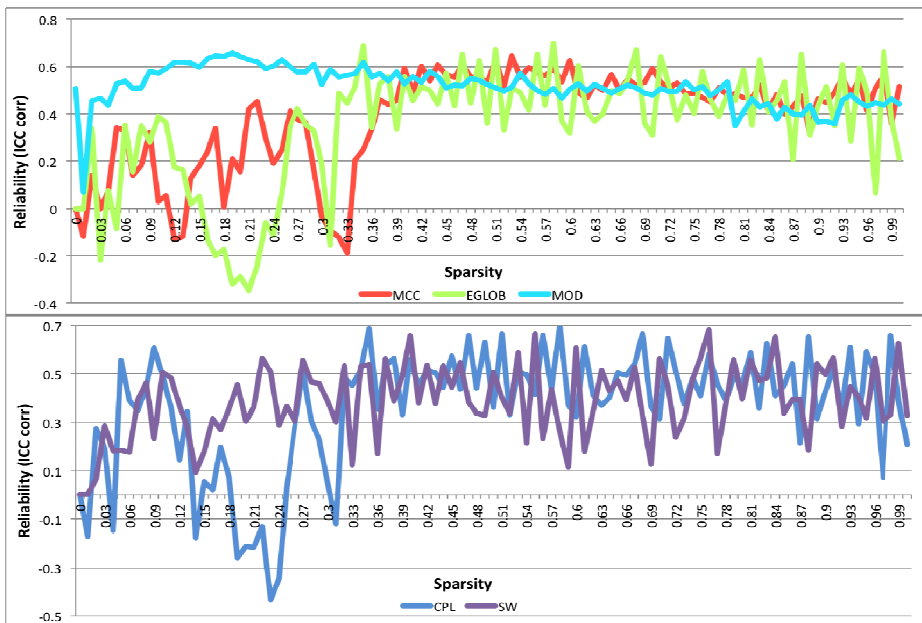
Fig. 2. Chart **of ICC r-values for 5 commonly used network topology measures across the full range of sparsities.** Clearly, the measures, and their reliability, depend on the sparsity threshold: this determines the proportion of connections retained, when sorted by edge strength. Retaining almost all connections (sparsity near 1) may include some that are unreliable, but using a very high threshold (sparsity near zero) may greatly affect which nodes are included in the network at all, promoting instability.

Next we assessed how reliable the global measures were when scores were integrated across a range of sparsities, to improve stability. The integral of the measures over the range 0.2-0.3 has been shown to be stable [4] and biologically plausible [20], yet we checked the test-retest reliability of scores integrated over 4 different ranges: 0.1-0.2, 0.2-0.3, 0.3-0.4, and 0.4-0.5. Values above 0.5 were not used because it is not considered biologically plausible [20]. Graph theory scores were integrated across these ranges, not the ICC *r*-values. Results for these test-retest reliability analyses are shown in Figure 3. For the 70x70 matrices, 57% of connections had a reliability of at least 0.30. The majority of the most reliable connections (≥0.70) were connections of the frontal cortex.
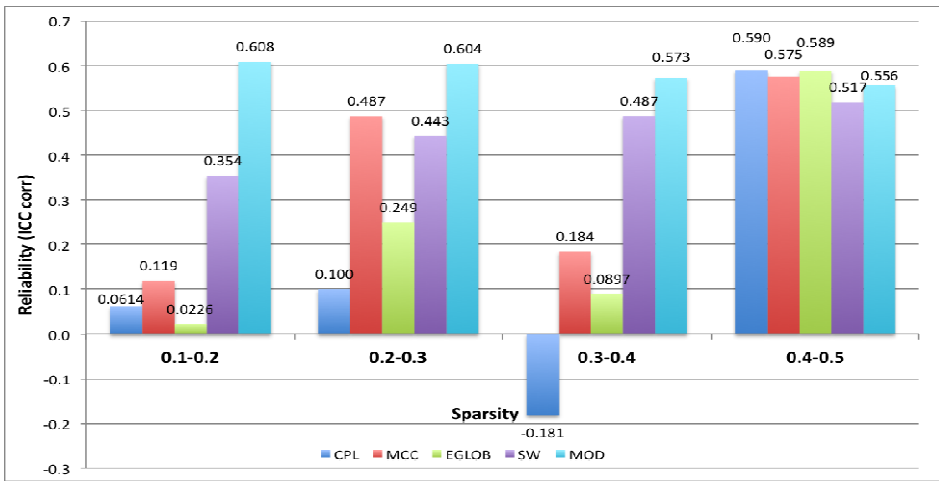
Fig. 3. ICC *r*-values show the reproducibility of 5 commonly used network measures, when they are integrated, to improve robustness, across 4 different sparsity ranges. MOD is still the most reliable measure. Also MOD is more stable than the other measures with respect to the sparsity, at least over the range examined here.

4   Discussion

In this paper we examined the test-retest reliability of graph theory measures of network connectivity applied to structural networks derived from HARDI scans. Reliability varied both across measures and across sparsities. Modularity was most reliable, and most stable with respect to sparsity threshold. This makes sense given that modularity has to do with how well the network can be broken into sub-networks – a measure of broad network topology, it may depend less on individual connections.

Characteristic path length and global efficiency are almost inverse measures of each other, except that global efficiency takes into account zeros in the connectivity matrix while characteristic path length does not. This could be responsible for the difference in reliability between characteristic path length and global efficiency. At low sparsities with more '0' entries, networks get fractured. Most measures are vulnerable to this loss of

nodes, but especially characteristic path length and global efficiency, as the mean shortest path length changes drastically if a significant portion of nodes is deleted. If networks get fractured differently between scan 1 and scan 2, this could lead to the very low reliability of characteristic path length and global efficiency at low sparsities. Characteristic path length and global efficiency are determined by calculating the path length between each node in a network and every other node in the network, for the shortest paths that exist, and averaging over all of those path lengths. Mean clustering coefficient, however, is determined by calculating for all the nodes connected to a given node, how many of its neighbors are also connected to each other, averaged over the whole network. Characteristic path length traces shortest paths, so if one path changes, many paths may take a different course, which could drastically alter mean shortest path length. For mean clustering coefficient, however, one path loss may reduce a node's clustering coefficient from 5/6 to 2/3 (for example); when averaged over the whole network this may not be a large net change.

Another factor that could be responsible for the difference in reliability between characteristic path length/global efficiency and mean clustering coefficient may be which paths are trimmed at low sparsities. Long-range paths heavily influence characteristic path length/global efficiency, but the mean clustering coefficient depends more on short paths. If long-range paths are generally trimmed before short-range paths, then the reliability of characteristic path length and global efficiency will drop sooner than that of mean clustering coefficient, as sparsity decreases, and their reliability will be impaired. In support of this, the reliability for characteristic path length, global efficiency, and mean clustering coefficient, are all much closer to each other at the highest sparsity, when all connections are retained.

There was a substantial dip in the reliability of a number of measures when integrated over the range 0.3-0.4. This was due to an increase in both the within- and between-subject variability in these measures. The average percent connectedness of these matrices was 26.5%, with all subjects fully connected at a sparsity of 0.30. The range of sparsities where all subjects are beginning to become fully connected may be associated with some instability in the measures, especially if many more unreliable (weak) connections are added.

5   Conclusion

Here we examined the test-retest reliability for a number of graph theory measures commonly used to assess brain structural connectivity. This depends to some extent on the tractography method, as well as the angular and spatial data resolution (we consider these topics elsewhere). Even so, we minimized several sources of error, using a 4-Tesla high angular resolution (94-direction) protocol, and a Hough method that uses ODFs to compute tracts. We found that modularity had moderately high reliability (mean r=0.58 for integrated analyses), as expected for a measure of general network topology. Mean clustering coefficient had higher reliability than characteristic path length or global efficiency for the lower sparsities, perhaps because networks fracture at lower sparsities. Integrating over a range of sparsities improved

the reliability of MCC and SW, while decreasing that of CPL and EGLOB at some sparsities. Selecting an appropriate sparsity range to integrate over, and defining network measures robust to sparsity, deserves further analysis.

References

1. Gong, G., et al.: Age- and gender-related differences in the cortical anatomical network. J. Neuroscience 29(50), 15684–15693 (2009)
2. Dennis, E., et al.: Heritability of structural brain connectivity network measures in 188 twins. In: SFN 2012, Washington, D.C., November 12-16 (2011)
3. Brown, J., et al.: Brain local network interconnectivity loss in aging APOE-4 allele carriers. PNAS 108(51), 20760–20765 (2011)
4. Dennis, E., et al.: Altered structural brain connectivity in healthy carriers of the autism risk gene, CNTNAP2. Brain Connectivity 1(6), 447–459 (2012)
5. Van den Heuvel, M., et al.: Aberrant frontal and temporal cortex network structure in schizophrenia: A graph theoretical analysis. J. Neuroscience 30(47), 15915–15926 (2010)
6. Bassett, D., et al.: Conserved and variable architecture of human white matter connectivity. NeuroImage 54(2), 1262–1279 (2011)
7. Wang, J.-H., et al.: Graph theoretical analysis of functional brain networks: Test-retest evaluation on short- and long-term resting-state functional MRI data. PLoS One 6(7), 1–22 (2011)
8. Telesford, Q., et al.: Reproducibility of graph metrics in fMRI Networks. Front. Neuroinform. 4, 1–10 (2010)
9. Deuker, L., et al.: Reproducibility of graph metrics in human brain functional networks. NeuroImage 47(4), 1460–1468 (2009)
10. Braun, U., et al.: Test-retest reliability of resting-state connectivity network characteristics using fMRI and graph theoretical measures. NeuroImage 59(2), 1404–1412 (2012)
11. Zalesky, A., et al.: Whole-brain anatomical networks: Does the choice of nodes matter? NeuroImage 50(3), 970–983 (2010)
12. Zhan, L., et al.: How do Spatial and angular resolution affect brain connectivity maps from Diffusion MRI? In: ISBI 2012, Barcelona, Spain, May 2-5 (2012)
13. Jones, D., et al.: Optimal strategies for measuring diffusion in anisotropic systems by magnetic resonance imaging. Magn. Res. Medicine 42(3), 515–525 (1999)
14. Holmes, C., et al.: Enhancement of MR images using registration for signal averaging. JCAT 22(2), 324–333 (1998)
15. Desikan, R., et al.: An automated labeling system for subdividing the hu-man cerebral cortex on MRI scans into gyral based regions of interest. NeuroImage 31, 968–980 (2006)
16. Aganj, I., et al.: Reconstruction of the orientation distribution function in single-and multiple-shell q-ball imaging within constant solid angle. Magn. Res. Medicine 64(2), 554–566 (2010)
17. Aganj, I., et al.: A Hough transform global probabilistic approach to mul-tiple-subject diffusion MRI tractography. Med. Image Anal. 15(4), 414–425 (2011)
18. Rubinov, M., Sporns, O.: Complex network measures of brain connectiv-ity: uses and interpretations. NeuroImage 52, 1059–1069 (2010)
19. Bullmore, E., Bassett, D.: Brain Graphs: Graphical Models of the Human Brain Connectome. Annu. Rev. Clin. Psychol., 1–37 (2010)
20. Sporns, O.: Networks of the Brain. MIT Press, Cambridge (2011)

Registration and Analysis of White Matter Group Differences with a Multi-fiber Model

Maxime Taquet[1,2], Benoît Scherrer[1], Olivier Commowick[3], Jurriaan Peters[1,4],
Mustafa Sahin[4], Benoît Macq[2], and Simon K. Warfield[1]

[1] Computational Radiology Laboratory, Children's Hospital Boston, Harvard, USA
[2] ICTEAM Institute, Université Catholique de Louvain, Louvain-La-Neuve, Belgium
[3] INRIA, INSERM, VisAGeS U746 Unit/Project, F-35042 Rennes, France
[4] Department of Neurology, Children's Hospital Boston, Harvard, USA

Abstract. Diffusion magnetic resonance imaging has been used extensively to probe the white matter in vivo. Typically, the raw diffusion images are used to reconstruct a diffusion tensor image (DTI). The incapacity of DTI to represent crossing fibers leaded to the development of more sophisticated diffusion models. Among them, multi-fiber models represent each fiber bundle independently, allowing the direct extraction of diffusion features for population analysis. However, no method exists to properly register multi-fiber models, seriously limiting their use in group comparisons. This paper presents a registration and atlas construction method for multi-fiber models. The validity of the registration is demonstrated on a dataset of 45 subjects, including both healthy and unhealthy subjects. Morphometry analysis and tract-based statistics are then carried out, proving that multi-fiber models registration is better at detecting white matter local differences than single tensor registration.

Keywords: Diffusion Imaging, Multi-Fiber Models, Registration, White Matter.

1 Introduction

Diffusion magnetic resonance imaging offers the ability to investigate in vivo the white matter microstructure. The representation of the signal by diffusion tensor images (DTI) has proven useful for population analysis in two ways [1]. First, scalar features extracted from DTI, such as the fractional anisotropy (FA), may indicate the presence of brain diseases. Second, the use of DTI in registration improves the detection of morphometric differences, compared to scalar images.

The single tensor diffusion model has, however, proven inaccurate for two main reasons. First, it cannot represent the signal arising from multiple fibers with heterogeneous orientations in one voxel. Second, it does not account for the non-monoexponential decay observed when imaging at high b-values. Novel models addressing one or both of these issues have been introduced [2] : Q-ball imaging, spherical deconvolution, 4th order tensors, DOT, and others. Most of them focus on describing the general shape of the diffusion at each voxel.

In contrast, mixture models represent each fiber bundle independently, keeping the interpretability of single fiber models while accounting for crossing fibers. Therefore, scalar quantities such as the fractional anisotropy (FA) can still be computed for each fiber independently. This property makes them very attractive for population analysis.

While the literature on the registration of complex diffusion models is growing (*e.g.*, [3,4]), no method has been developed to register mixture models. This lack of a registration method limits the use of mixture models in population analysis despite their attractiveness. This issue has been previously reported, and has incited researches to register the raw diffusion weighted images instead [5].

The remaining of this paper is organized as follows. Section 2 introduces the diffusion mixture models. Section 3 presents a method to compute weighted average of mixture models. Section 4 develops a similarity metric for diffusion mixture images. Section 5 presents the integration of the developed methods in a registration algorithm and analyzes its complexity. Section 6 presents experimental results on a dataset of 45 subjects. Finally, Section 7 concludes.

2   Diffusion Mixtures

The basic idea behind multi-fiber models is to fit a single fiber model to each of the fiber bundles present in the voxel. If $S_i(\boldsymbol{x})$ is a suitable model to represent the diffusion process in a single fiber, then,

$$S(\boldsymbol{x}) = \sum_{i=1}^{N} f_i S_i(\boldsymbol{x}) \tag{1}$$

is a multi-fiber model for N crossing fibers with relative volumetric occupancy given by f_i. The assumption behind these models is that the exchange of water molecules between populations of fibers is negligible during the diffusion time [2].

The simplest multi-fiber model is the multi-tensor model in which $S_i = S_0 e^{-b\boldsymbol{g}^T \boldsymbol{D}_i \boldsymbol{g}}$. More complex multi-fiber models have later been introduced [6]. Potentially, any single fiber model can be extended to a multi-fiber model by means of mixtures. One such model, the biexponential decay model [2], represents each fiber bundle by a Gaussian mixture to capture the non-monoexponential decay of the signal. The corresponding multi-fiber model would be a mixture of Gaussian mixtures which is itself a Gaussian mixture. A natural parameterization of diffusion Gaussian mixtures is the set of pairs (fraction, covariance matrix), that we write: $\{(f_1, \boldsymbol{\Sigma}_1), ..., (f_N, \boldsymbol{\Sigma}_N)\}$. Alternatively, to connect with the tensor formalism, the inverse of the covariance matrix, $\boldsymbol{D}_i = \boldsymbol{\Sigma}_i^{-1}$, can be used.

3   Weighted Combination of Mixtures

Computing weighted combinations of voxel values is at the basics of interpolation (the value in one location is the weighted combination of the values in the

neighborhood), smoothing (the value at a grid voxel is replaced by a weighted combination of the values in a neighborhood) and atlas construction (the value at one voxel is the average of the values in the aligned subjects').

Gaussian mixture simplification (GMS) was introduced to efficiently compute weighted combinations of diffusion mixture [7]. In this section, we underline the important aspects of this method. The idea behind GMS is that computing weighted combinations of mixtures would be trivial if the number of components of the result could be arbitrarily large. Indeed, the linear combination of K mixture models with N components is a mixture models with $M = KN$ components:

$$\mathcal{M}_C = \sum_{k=1}^{K} w_k \mathcal{M}_k = \sum_{k=1}^{K} w_k \sum_{j=1}^{N} f_j^k S_j^k(\boldsymbol{x}) \equiv \sum_{i=1}^{M} g_i S_i(\boldsymbol{x}). \tag{2}$$

We refer to this mixture as the *complete* mixture. GMS optimizes the parameters of a *simplified mixture* $\mathcal{M}_S = \sum_{j=1}^{N} h_j R_j(\boldsymbol{x})$ with $N < M$ components that best approximates $\mathcal{M}_C$. The energy function to be minimized is the cumulative differential entropy (the reference to the location $\boldsymbol{x}$ is omitted for clarity):

$$D(\mathcal{M}_C, \mathcal{M}_S) = \sum_{j=1}^{N} \sum_{i:\pi_i=j} g_i D(S_i||R_j) = \sum_{j=1}^{N} \sum_{i:\pi_i=j} g_i \int S_i(\boldsymbol{g}) \log \frac{S_i(\boldsymbol{g})}{R_j(\boldsymbol{g})} d\boldsymbol{g}, \tag{3}$$

where $\boldsymbol{g}$ is the gradient vector and where latent variables π_i cluster the components of the complete mixture S_i in N clusters each represented by a single component of the simplified mixture, R_j; $\pi_i = j$ means that S_i is best represented by R_j. Following the recent developments in probabilistic clustering, an EM scheme is used to minimize (3). Banerjee *et al* showed that both the E-step and the M-step can be solved in closed form for mixtures of exponential distributions [8]. For Gaussian mixtures, the E-step consists in optimizing the latent variables π_i by computing the Burg matrix divergence between the covariance matrices of each component of $\mathcal{M}_C$ ($\boldsymbol{\Sigma}_i^S$) and each component of $\mathcal{M}_S$ ($\boldsymbol{\Sigma}_j^R$):

$$\pi_i = \arg\min_j B\left(\boldsymbol{\Sigma}_i^S, \boldsymbol{\Sigma}_j^R\right) = \arg\min_j \text{Tr}\left(\boldsymbol{\Sigma}_i^S \boldsymbol{\Sigma}_j^{R-1}\right) - \log\left|\boldsymbol{\Sigma}_i^S \boldsymbol{\Sigma}_j^{R-1}\right|. \tag{4}$$

As for the M-step, it sums up to calculating:

$$\boldsymbol{\Sigma}_j^R = \frac{\sum_{i:\pi_i=j} f_i \boldsymbol{\Sigma}_i^S}{\sum_{i:\pi_i=j} f_i} \qquad \text{and} \qquad h_j = \sum_{i:\pi_i=j} f_i. \tag{5}$$

Alternating (4) and (5) until convergence provides the parameters (h_j and $\boldsymbol{\Sigma}_j^R$) of the resulting mixture. A log-Euclidean version of this interpolation scheme is obtained by replacing all covariance matrices by their logarithm.

4   Generalized Correlation Coefficient for Mixtures

The correlation coefficient, invariant under linear transformations of the voxel intensities, is widely used in mono-modal image registration. The inter-subject
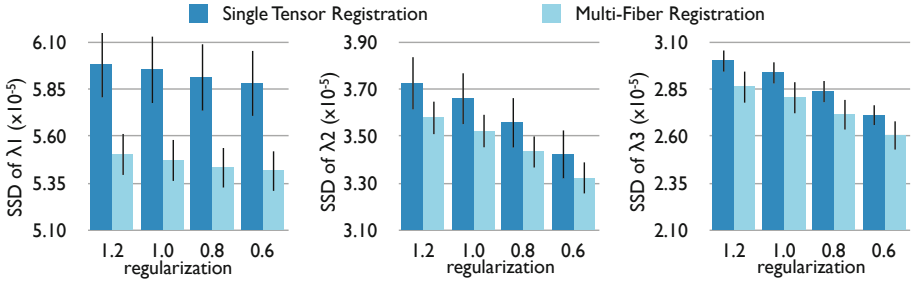
Fig. 1. Comparison of the single tensor and multi-fiber registration in terms of the SSD between eigenvalues after alignment, for different regularization parameter values [9]. Multi-fiber registration significantly improves the quality of the registration.

variability of diffusivity values motivates the introduction of a generalized correlation coefficient, invariant under these differences. In DTI, this variability has been reported and partially accounted for in some registration methods [1]. The correlation coefficient between blocks F and G is defined as the scalar product of the normalized blocks:

$$\rho(F, G) = \left\langle \frac{F - \mu_F}{||F - \mu_F||}, \frac{G - \mu_G}{||G - \mu_G||} \right\rangle,$$

where μ_F is the mean of the image values in the block. It is invariant if F (and/or G) is replaced by $aF + b$. It has been generalized to vector images by redefining the means μ_F and μ_G as the projection of the block onto a constant block T [10]:

$$F - \mu_F = F - \langle F, T \rangle \frac{T}{||T||^2}.$$

The corresponding generalized correlation coefficient is invariant if F is replaced by $aF + bT$ where T is now any constant vector block. The definition of a scalar product between two blocks of mixture models seems impractical if not impossible. We therefore further generalize the correlation coefficient by substituting the inner product by a more general scalar mapping, $m(\mathcal{M}_f, \mathcal{M}_g)$:

$$\rho(\mathcal{M}_f, \mathcal{M}_g) = m\left(\frac{\mathcal{M}_f - m(\mathcal{M}_f, \mathcal{T})\mathcal{T}}{n_m(\mathcal{M}_f - m(\mathcal{M}_f, \mathcal{T})\mathcal{T})}, \frac{\mathcal{M}_g - m(\mathcal{M}_g, \mathcal{T})\mathcal{T}}{n_m(\mathcal{M}_g - m(\mathcal{M}_g, \mathcal{T})\mathcal{T})} \right),$$

where $n_m(\mathcal{M})^2 = m(\mathcal{M}, \mathcal{M})$ is a generalization of the norm. This definition does not guarantee the invariance property of the metric for any scalar mapping. One can show that the invariance is preserved as long as the scalar mapping is linear with respect to the constant block $\mathcal{T}$:

$$m(a\mathcal{M}_f + b\mathcal{T}, \mathcal{T}) = a\, m(\mathcal{M}_f, \mathcal{T}) + b\, m(\mathcal{T}, \mathcal{T}). \tag{6}$$

To preserve the interpretability of ρ as a similarity metric, it needs to be symmetric, equal to one in case of perfect match and lower than one in any other case. These constraints on ρ translate into the following constraints on m:
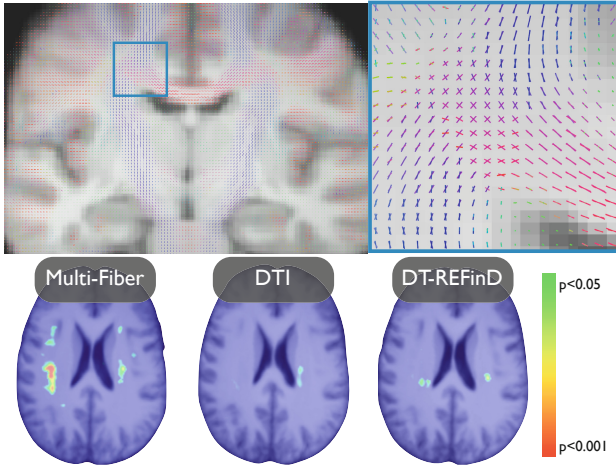
Fig. 2. (top) The two-tensor atlas built by means of the developed registration method reveals crossing pathways common to all anatomies. (bottom) White matter volume shrinkage in tuberous sclerosis represented by the p-value maps. Multi-fiber registration reveals more differences than single tensor registration and DT-REFinD [11]

$$m(\mathcal{M}_f, \mathcal{M}_g) = m(\mathcal{M}_g, \mathcal{M}_f) \tag{7}$$

$$n_m(a\mathcal{M}_f) = a\, n_m(\mathcal{M}_f) \tag{8}$$

$$|m(\mathcal{M}_f, \mathcal{M}_g)| \leq n_m(\mathcal{M}_f) n_m(\mathcal{M}_g). \tag{9}$$

The latter is a generalized form of the Cauchy-Schwartz inequality for inner products. Conditions (6-9), the choice of $\mathcal{T}$ and the definitions of the addition $\mathcal{M} + \mathcal{T}$ and multiplication by a scalar $a\mathcal{M}$, stand together as a model to define a correlation coefficients in potentially any space. For DTI, if $\mathcal{T}$ is an isotropic tensor block ($\mathcal{T}(\boldsymbol{x}) = D\boldsymbol{I}_{3\times3}$), m is the log-Euclidean scalar product, and the log-Euclidean algebra is used, then ρ is invariant under linear transformations of the eigenvalues in the log-domain [12]. For multi-tensor images, we fix $\mathcal{T}(\boldsymbol{x}) = \left\{ (\frac{1}{N}, D\boldsymbol{I}_{3\times3}), ..., (\frac{1}{N}, D\boldsymbol{I}_{3\times3}) \right\}$, and we define the addition of $\mathcal{T}$, and the multiplication by a scalar component-wise in the log-domain. The scalar mapping $m(\mathcal{M}_f, \mathcal{M}_g)$ is defined by pairing the tensors in each voxel to maximize the linear combination of pairwise scalar products. Let $\mathcal{M}_f(\boldsymbol{x}) = \{(f_1, \boldsymbol{F}_1), ..., (f_N, \boldsymbol{F}_N)\}$ and $\mathcal{M}_g(\boldsymbol{x}) = \{(g_1, \boldsymbol{G}_1), ..., (g_N, \boldsymbol{G}_N)\}$ defined on a domain Ω, we have:

$$m(\mathcal{M}_f, \mathcal{M}_g) = \sum_{\boldsymbol{x}\in\Omega} \max_{\pi} \sum_{i=1}^{N} f_i g_{\pi(i)} \left\langle \boldsymbol{F}_i, \boldsymbol{G}_{\pi(i)} \right\rangle,$$

where π is a pairing function associating one tensor of $\mathcal{M}_g$ to each tensor of $\mathcal{M}_f$. This scalar mapping satisfies conditions (6-9). Interestingly, the resulting generalized correlation coefficient is invariant under any global (within the block) linear transformation of all eigenvalues in the log-domain. This similarity metric is therefore robust to the inter-subject variability of diffusivities.
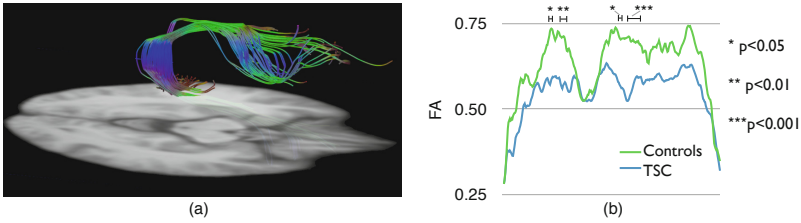
Fig. 3. (a) Arcuate fasciculus, a set of fibers involved in language, on which tract based statistics was performed, (b) The FA profile in TSC patients shows significantly disrupted white matter fascicules in different clusters, indicated by the stars.

5   Implementation and Complexity

The developed methods were integrated in the efficient block matching registration algorithm described in [9]. The parameters used are the following: 4 pyramid level, 10 iterations per level, block size: $5 \times 5 \times 5$, outlier removal rate: 20%. The implementation was multi-threaded. On a 8 core workstation, with $220 \times 220 \times 176$ two-fiber images, the entire registration takes 1.5 hour. All weighted combinations were computed until complete convergence of the soft clustering. The average number of iterations required for that convergence is 4.

6   Results

The registration was applied to a clinical dataset of 45 subjects, 13 controls and 32 patients with tuberous sclerosis complex (TSC), a rare genetic disease associated with impaired white matter integrity. A DTI and a multi-tensor model with three components (one isotropic and two anisotropic) were reconstructed for each subject [13].

6.1   Validation

An alternative to the method presented in this paper would be to select one of the two tensors in each voxel (e.g. the one with the highest FA) and to perform single tensor registration on this image. Here, we validate that our method works better than this simple alternative. The quality of the alignment is assessed by the sum of square differences of each eigenvalue after alignment of control subjects. Indeed, while the diffusivities can significantly differ in diseased brain, they are approximately equal for healthy subjects. We performed 26 randomly chosen registrations with four levels of regularization, totalizing 104 registrations. In each voxel, the eigenvalues were averaged between the two anisotropic components (weighted by their fractions). Results show that multi-fiber registration performs significantly better than single tensor registration (Fig. 1).

6.2   Atlas Construction

An atlas was constructed using our registration and alternating three steps: aligning all subjects to the current atlas (initially a randomly chosen subject), averaging the aligned subjects (using the weighted combination of mixtures), applying the mean inverse field to the resulting average [14]. This atlas remarkably shows areas where multiple fibers are consistently present in all subjects (Fig. 2).

6.3   Morphometry

The clinical hypothesis according to which there is substantial white matter shrinkage in TSC subjects was tested by performing a one-tailed two sample t-test on the log-Jacobian of the deformation fields [15]. The subject classes were then randomly permuted 4000 times to assess the null distribution of extreme t-scores. The entire process was repeated with single tensor images. As a result, multi-fiber registration reveals more white matter differences (> 3800 significant voxels) than single tensor (< 1000 voxels) (Fig 2). The entire process was then repeated with DT-REFinD, a state-of-the art DTI registration algorithm [11], to test whether the improved detection of differences is truly due to the knowledge brought by multi-fiber models. Again, DT-REFinD did not capture all the differences detected by multi-fiber registration (< 1300 voxels) (Fig 2).

6.4   Tract-Based Statistics

Some structural subnetworks are believed to be impaired in TSC patients. To test this hypothesis, we analyzed the FA profile along the median tract of the arcuate fasciculus, generated on the atlas by a probabilistic tractography algorithm [5] (Fig. 3(a)). A one-tailed two-sample t-test was performed at every location. A threshold t_0 was then set to the t-statistics and the length of the contiguous supra-threshold segments were recorded. The null distribution of these lengths was assessed by randomly permuting the subjects classes 4000 times. The operation was repeated for a wide range of thresholds ($1.5 \leq t_0 \leq 4.5$) to estimate the robustness of our findings. For $t_0 = 2.7$ ($p_0 = 0.01$), with our multi-fiber registration, four significant clusters, together representing 15% of the tract were detected, indicating a strong impairment of this subnetwork in TSC patients (Fig. 3(b)). These findings were robust to the choice of t_0 for any $t_0 \leq 3.4$. In contrast, single tensor registration only revealed one cluster representing 5% of the fiber, which was not robust outside the range $1.9 \leq t_0 \leq 2.8$.

7   Conclusions

This paper introduced a registration and atlas construction method to align multi-fiber models. A proper interpolation method and a robust similarity metric were presented. Results in both morphometry and tract-based statistics demonstrated that multi-fiber registration reveals more group differences than DTI

registration. We therefore believe that this registration method opens new doors to understanding brain disorders based on multi-fiber models.

References

1. Zhang, H., Avants, B., Yushkevich, P., Woo, J., Wang, S., McCluskey, L., Elman, L., Melhem, E., Gee, J.: High-dimensional spatial normalization of diffusion tensor images improves the detection of white matter differences: an example study using amyotrophic lateral sclerosis. IEEE TMI 26(11), 1585–1597 (2007)
2. Minati, L., Weglarz, W.: Physical foundations, models, and methods of diffusion magnetic resonance imaging of the brain: A review. Concepts in Magnetic Resonance Part A 30(5), 278–307 (2007)
3. Barmpoutis, A., Vemuri, B.C., Forder, J.R.: Registration of High Angular Resolution Diffusion MRI Images Using 4th Order Tensors. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part I. LNCS, vol. 4791, pp. 908–915. Springer, Heidelberg (2007)
4. Yap, P., Chen, Y., An, H., Yang, Y., Gilmore, J., Lin, W., Shen, D.: Sphere: Spherical harmonic elastic registration of hardi data. NeuroImage 55(2), 545–556 (2011)
5. Bergmann, O., Kindlmann, G., Peled, S., Westin, C.: Two-tensor fiber tractography. In: IEEE International Symposium on Biomedical Imaging, pp. 796–799 (2007)
6. Assaf, Y., Basser, P.: Composite hindered and restricted model of diffusion (charmed) mr imaging of the human brain. Neuroimage 27(1), 48–58 (2005)
7. Taquet, M., Scherrer, B., Benjamin, C., Prabhu, S., Macq, B., Warfield, S.: Interpolating multi-fiber models by gaussian mixture simplification. In: IEEE International Symposium on Biomedical Imaging (2012)
8. Banerjee, A., Merugu, S., Dhillon, I., Ghosh, J.: Clustering with bregman divergences. The Journal of Machine Learning Research 6, 1705–1749 (2005)
9. Commowick, O., Arsigny, V., Isambert, A., Costa, J., Dhermain, F., Bidault, F., Bondiau, P., Ayache, N., Malandain, G.: An efficient locally affine framework for the smooth registration of anatomical structures. MedIA 12(4), 427–441 (2008)
10. Ruiz-Alzola, J., Westin, C., Warfield, S., Alberola, C., Maier, S., Kikinis, R.: Nonrigid registration of 3d tensor medical data. MedIA 6(2), 143–161 (2002)
11. Yeo, B., Vercauteren, T., Fillard, P., Peyrat, J., Pennec, X., Golland, P., Ayache, N., Clatz, O.: Dt-refind: Diffusion tensor registration with exact finite-strain differential. IEEE Trans. on Medical Imaging 28(12), 1914–1928 (2009)
12. Taquet, M., Macq, B., Warfield, S.: A generalized correlation coefficient: Application to dti and multi-fiber dti. In: IEEE MMBIA, pp. 9–14 (2012)
13. Scherrer, B., Warfield, S.: Toward an accurate multi-fiber assessment strategy for clinical practice. In: IEEE International Symposium on Biomedical Imaging, pp. 2140–2143 (2011)
14. Guimond, A., Meunier, J., Thirion, J.: Average brain models: A convergence study. Computer Vision and Image Understanding 77(2), 192–210 (2000)
15. Ashburner, J., Hutton, C., Frackowiak, R., Johnsrude, I., Price, C., Friston, K.: Identifying global anatomical differences: deformation-based morphometry. Human Brain Mapping 6(5-6), 348–357 (1998)

Scalable Tracing of Electron Micrographs by Fusing Top Down and Bottom Up Cues Using Hypergraph Diffusion

Vignesh Jagadeesh, Min-Chi Shih, B.S. Manjunath, and Kenneth Rose⋆

Department of ECE and Center for Bioimage Informatics
University of California, Santa Barbara CA - 93106

Abstract. A novel framework for robust 3D tracing in Electron Micrographs is presented. The proposed framework is built using ideas from hypergraph diffusion, and achieves two main objectives. Firstly, the approach scales to trace hundreds of targets without noticeable increase in runtime complexity. Secondly, the framework yields flexibility to fuse top down (global cues as hyperedges) and bottom up (local superpixels as nodes) information. Subsequently, a procedure for auto-seeding to initialize the tracing procedure is proposed. The paper concludes with experimental validation on a challenging large scale tracing problem for simultaneously tracing 95 structures, illustrating applicability of the proposed algorithm.

Keywords: Tracking, Tracing, Electron Micrograph, Spectral Graphs.

1 Introduction

Connectomics [1] is a sub-field of bio-informatics attempting to understand neuronal connectivity patterns from data acquired using microscopic imaging of neurons. This work focusses on the analysis of volumetric datasets from a connectome, acquired using electron microscopy at nanometer resolutions. The major image analysis challenge in tracing neuronal structures from Electron Micrographs (EM) are two-fold. Firstly, the datasets are extremely large with a requirement to scale algorithms to trace hundreds of targets (structures) simultaneously. Secondly, the structures present in the data undergo arbitrary deformations and topological changes that need to be accurately modeled. This work proposes a tracing model attempting to jointly satisfy the above requirements. The problem can be defined as one of extracting 3D reconstructions of hundreds of structures from a volumetric dataset in an accurate and computationally efficient manner. There are well established algorithms for single structure tracing in Electron Micrographs with deformations and topological changes. However, creating multiple binary segmentations by applying single structure tracing on

every structure is problematic. Firstly, if a pixel is set to one in the binary masks of two structures, it is not clear which label takes ownership of the structure. Secondly, interactions between structures cannot be modeled in such a scenario. The obvious solutions of utilizing discrete Markov random fields(MRFs)/Level sets, though attractive may not be most suitable as verified by experiments in Section 3. In case of MRFs, scaling the number of labels has a direct impact on the runtime of algorithms like alpha expansions and behavior of such methods for segmenting hundreds of labels is not a well studied problem (though such problems have been looked into for stereo and optic flow). Level Set methods like the Chan-Vese model, have also not been shown to work on hundreds of labels.

Proposed Solution: The image stack is assumed to be made up of superpixels linked to each other in three dimensions, forming a graph. As an example, Figure 1a shows superpixel segmentation of two consecutive slices, say s13 represents third superpixel in slice 1. The superpixel graph is constructed by introducing edges between the superpixel of interest (s11) and its spatial (s12, s14) and temporal neighbors (s21, s22, s24). Hyperedge construction from top down information is illustrated in Figure 1b. Maroon dotted circle is the output of top down detector grouping red, green, blue and yellow blobs leading to a hyperedge. Hyperedges based on k-nearest neighbors are similarly constructed. The key idea is to model the label propagation across image sequences as the solution of a hypergraph diffusion equation in a 3D superpixel hypergraph. In doing so, one is presented with a model having some very desirable properties. Firstly, a flexible framework that can utilize top down (coarse object location) and bottom up (local image structure) information results. Secondly, the diffusion has a closed form solution with complexity weakly dependent on the number of labels. In
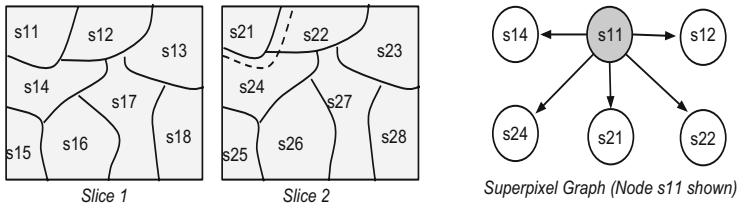
(a) Three Dimensional Superpixel Graph Construction
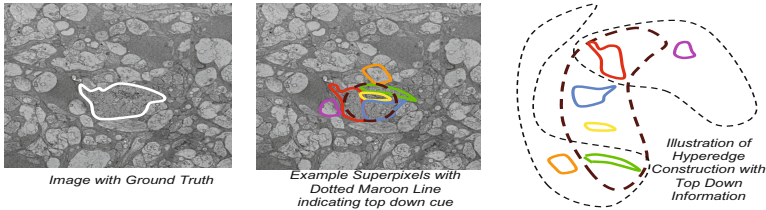
(b) Hyperedge Construction from Superpixels

Fig. 1. Construction of the 3D superpixel hypergraph

other words, the complexity remains unchanged in spite of an increase in the number of labels. The primary contributions of this work include: ***An efficient tracing framework based on hypergraph diffusion that fuses top down and bottom up cues*** and a method for ***Automatic Target Seeding***.

Related Works: The works by [7,8,6,12] are good sources of reference for EM image analysis. Further, [5] utilized hypergraphs for unsupervised video segmentation, in contrast to the supervised case the proposed approach deals with. Salient aspects of the proposed 3D tracing framework are scalability to hundreds of labels, modeling higher order interaction between segments, introduction of global contour cues using hyperedges, generic autoseeding for fully automatic tracing, and semi-supervised nature, amenable to user interaction if needed. Our claim of originality is in the framework comprising the salient aspects listed above. In related work, [12] propose an interesting technique utilizing pairwise segment interactions on EM data from a mouse (gradient based), but do not lay emphasis on user interaction or scalability. The data used in this work is very different and is from a rabbit retina (noisy regional texture based). Techniques similar to [12] did not perform well on our datasets, leading us to compare with the state of the art on rabbit retina and relevant tracing techniques. Our work is intended as a scalable replacement to the graph cut solvers used in [6], as will be established by experimental results.

2   Proposed Model

The model is initially presented in terms of a bigraph. Subsequently, it is extended to the hypergraph case. The intuition behind the graph diffusion energy is motivated by semi-supervised learning [13], where structure of the data manifold is utilized along with a sparse initial labeling of data points (y) to arrive at a final labeling(f). Consider a bigraph $G = (V, E, w)$, comprising a vertex set $v \in V$ with weights between nodes $\{u, v\} \in G$ denoted by $w(u, v)$. Further, let $d(u)$ denote the degree of node u. Considering a two label model, $f \in \{-1, 1\}^{|V|}$ is the classification function to be estimated, and $y \in \{-1, 1\}^{|V|}$ is the initial labeling vector. The following equation can be interpreted as follows, estimate a labeling function f over graph G whose smoothness is measured by a smoothness cost term, and which does not deviate too much from initial

labeling: $\underset{f \in R^{|V|}}{\operatorname{argmin}} \frac{1}{2} \underbrace{\sum_{\{u,v\} \in V} w(u,v) \left(\frac{f(u)}{\sqrt{d(u)}} - \frac{f(v)}{\sqrt{d(v)}} \right)^2}_{Smoothness\ Cost} + \underbrace{\mu ||f - y||^2}_{Deviation\ from\ Seeds}$.

The above formulation can be extended to a hypergraph [14], generalizing the notion of an edge linking a pair of nodes to a hyperedge linking multiple nodes. The intuition behind the hypergraph model is similar to the bigraph case, except for the fact that smoothness is over multiple nodes constituting a hyperedge. Consider a hypergraph $G = (V, E, w)$, comprising a vertex set $v \in V$,

an edge set $e \in E$ and a set of weights w. A hyperedge e comprises of a set of nodes $\mathbf{v_e} \subset V$ that form a clique inside the hyperedge. The degree of a hyperedge is $\delta(e) = |\mathbf{v_e}|$, while the degree of a vertex is defined by $d(v) = \sum_{e \in E, v \cap \mathbf{v_e} \neq \varnothing} w(e)$. The incidence matrix $H \in \Re^{|V| \times |E|}$ contains binary elements h(v,e) taking the value 1 if $v \in v_e$, and 0 otherwise. $D_e \in \Re^{|E| \times |E|}$ and $D_v \in \Re^{|V| \times |V|}$ refer to the diagonal matrices of hyperedge and vertex degrees. The ultimate goal is to perform estimation of a smooth function f on the graph, given an initial labeling $y \in \{-1, 1\}^{|V|}$. The formulation to accomplish the same is

$$\text{given by: } \underset{f \in R^{|V|}}{\operatorname{argmin}} \underbrace{\frac{1}{2} \sum_{e \in E} \sum_{\{u,v\} \subset e} \frac{w(e)}{\delta(e)} \left(\frac{f(u)}{\sqrt{d(u)}} - \frac{f(v)}{\sqrt{d(v)}} \right)^2}_{Smoothness\ Cost} + \underbrace{\mu||f - y||^2}_{Deviation\ from\ Seeds} .$$

Defining the matrix $\Theta = D_v^{-\frac{1}{2}} H W D_e^{-1} H^T D_v^{-\frac{1}{2}}$, and $\triangle = I - \Theta$, it is straightforward to obtain a closed form solution on label certainties by, $f = (1 - \zeta)(I - \zeta\Theta)^{-1}y, \zeta = \frac{1}{1+\mu}$. In the context of image labeling (see Figure 1), $v \in V$ corresponds to a set of superpixels in an image sequence, $e \in E$ refer to the hyperedges constructed by including higher order neighbors on the 3D superpixel graph, thus forming the matrix H. Inferring using the above equation would result in a class marginal on each superpixel that would lead to a tracing result. The use of transductive hypergraph learning for supervised tracing in EM stacks is our first contribution.

Extension to Multiple Labels and Uncertainty Characterization: The above formulation can be extended to the multiple label case in a straightforward manner. The vectors f, y used for the two class problems are now transformed to matrices $F, Y \in \Re^{|V| \times |\mathcal{L}|}$, where column j of $\{F, Y\}$ correspond to the probability of label j to be associated with every node in the graph. The entry $Y(i,j)$ is set to 1 if node i has a label j associated with it, and $F(i,j)$ yields the probability of node i to be associated with label j after diffusion. Alternately, each row i of the matrices F, Y can be interpreted as the probabilities of node i to be associated with each label. The associated inference is given by: $F = (1 - \zeta)(I - \zeta\Theta)^{-1}Y, \zeta = \frac{1}{1+\mu}$. A side benefit of the above formulation is the fact that uncertainty of solutions can be characterized from the entropy computed using rows of F. Computing uncertainty estimates would point towards confidence of the algorithm in its solutions, and it can readily probe the user for assistance using an active learner in interactive settings.

Low Level Features and Graph Weights: The feature representation of superpixels plays an important role in the end results. We utilize gray scale and Local Binary Pattern (LBP) based texture histograms [9] for characterizing appearance of superpixels. The distance between histograms is modeled using the symmetric Kullback Leibler divergence, assuming independence between gray scale and texture channels. Assuming $h_{gray}(i)$ and $h_{lbp}(i)$ respectively to denote the gray scale and texture histograms of superpixel i, the dissimilarity

between superpixel i and j is constructed as: $KL[h(i), h(j)] = \sum_k h(i,k) \ln \frac{h(i,k)}{h(j,k)} + h(j,k) \ln \frac{h(j,k)}{h(i,k)}$ $w(i,j) = \exp\left(-KL[h_{gray}(i), h_{gray}(j)] + KL[h_{lbp}(i), h_{lbp}(j)]\right)$.

Complexity of Algorithm. The complexity of inversion is cubic $\mathcal{O}(|V|^3)$ in the number of nodes, as is evident from the equation for inferring multiple labels. Since the matrix considered is sparse, efficient sparse solvers can be employed leading to considerable reduction in running time. The matrix inversion of the graph Laplacian has the greatest computational load, while the matrix multiplication with the label vector Y is of lower complexity than the inversion. As a result, an increase in the columns of Y (additional targets) does not affect the overall time complexity of the algorithm. An intuitive way of looking at the solution is that the graph Laplacian models the entire 3D stack (primary target and contextual information), and a diffusion utilizing this graph Laplacian yields a simple and efficient method for label propagation. Finally, if solutions need to be corrected during interactive segmentation, the computed inverse matrix can be cached, resulting in extremely fast responses to user corrections. Now two important questions arise, Can the hyperedges be utilized to induce a top down global contour cue? (Global Cue Detectors) and is it possible to automatically initialize the number of targets present in the field of view? (Automatic Seeding). The following discussions answer the above questions followed by experimental validation of the proposed ideas.

Global Cue Detectors. As has been described, global detectors are outputs of any algorithm that gives a rough grouping of the nodes in a graph. In the current problem, any algorithm that gives a probable association between superpixels over the 3D volume, thus modeling higher order correlation over the stack is called a global cue detector. The idea is to learn edge profiles using Boosted Edge Learning (BEL) [4], followed by a pass of watershed transform for obtaining 2D segments. The 2D segments are associated in an unsupervised manner across the third dimension using the Floyd-Warshall all source shortest path algorithm [3] to generate probable global cues, see Figure 2a. These cues define association rules between superpixels, thus modeling longer range correlations. Subsequently, the k-nearest neighbors of every superpixel are also used in constructing hyperedges constituting additional global cues.

An important observation to be made is that the k-nearest neighbor hyperedges enforce a Potts style prior encouraging spatial smoothness of labels among superpixels using pairwise interactions. However, in scenarios where information on association between superpixels over larger spatial neighborhoods (across space and the third dimension) are available, they can be readily encoded using the hyperedges for promoting label smoothness. The top down cue detectors serve the purpose of defining these larger interaction neighborhoods over which label smoothness is encouraged.

Automatic Seeding. Another important problem that arises is the automatic initialization (seeding) of targets for efficient tracing. While the most widely used strategy for target initialization is based on user marked seeds in the first

(a)   (b)
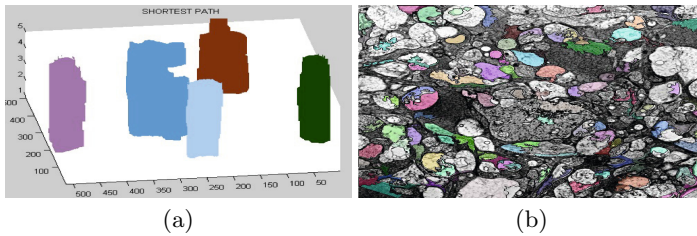
Fig. 2. Figure of the left illustrates outputs from the top down cue detector that serves to construct hyperedges. Figure on right illustrates results of the automatic seeding.

frame, it may not always be possible to seed hundreds of targets manually. In order to solve the above problem, a technique for automatic seeding is proposed. The result of the seeding algorithm can be utilized for initializing the matrix Y. The question asked for auto-seeding is: *Is it possible to pick a set (cardinality $|\mathcal{L}|$) from the K superpixels from the first slice that are as different from one another in appearance $(A_{i1} \in \Re^m)$ and spatial positions $(S_{i1} \in \Re^2)$?* The above problem can be solved by estimating an indicator vector $z \in \{0,1\}^K$ that minimizes a cost comprising distances between selected points in spatial and feature spaces. We formalize the above intuition as a relaxed quadratic program [2] (QP) $\operatorname*{argmin}_{z} \sum_{i=1}^{|V|} \sum_{j=1}^{|V|} w_{ij} z_i z_j, s.t \quad \sum_{i=1}^{N} z_i = |\mathcal{L}|$. Rounding the solution of QP yields the desired set of superpixels to be used as seeds for tracing, see Figure 2b. The weights w in the above equation can be constructed in a manner discussed previously.

Contour Refinement. The result of hypergraph diffusion achieves regional homogeneity but is not always edge aware. We utilize an edge based active contour based on hidden Markov models (HMM) [10][11]. Contours resulting from hypergraph diffusion initialize the edge based active contour. For each contour, a trellis with states sampled as points along normals to the contour is instantiated. These points represent the states of the HMM, and any path through the trellis is a potential contour candidate. The Viterbi decoding algorithm yields the final contour passing through strong image gradients.

3   Experiments

Experiments are reported on a tracing task over two separate stacks of electron micrographs. For the purpose of studying the behavior of diffusion in isolation from seeding, contours are manually initialized in the first frame. The metrics used for validating the tracing are the F-measure and Rand index, two commonly used metrics in the segmentation literature. Justification for scalability is given by the running time of algorithms on an Intel Core i7 860 @ 2.8GHz machine. Further, two variants of the proposed idea are utilized in experiments. The HGraph3D method attempts to perform diffusion through the entire 3D graph
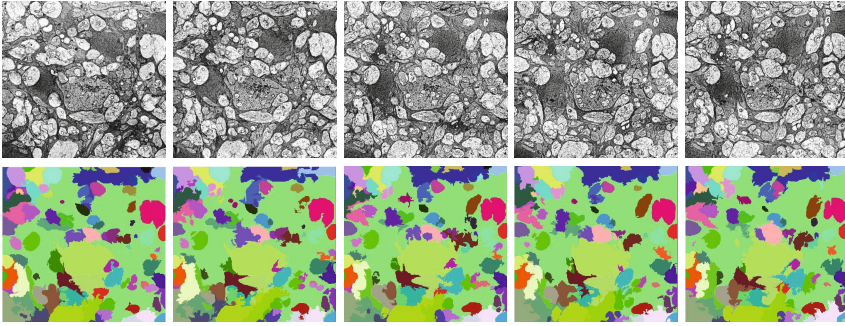
(a) Result of hypergraph diffusion on Dataset-I

Method	Median	Mean	Standard Deviation	Time(sec.)
Level Set Tracking	0.55	0.53	0.25	1080
BiGraph3D	0.16	0.23	0.26	78
HGraph3D (Proposed)	0.68	0.66	**0.21**	108
HGraph Propagate (Proposed)	**0.71**	**0.78**	0.22	**34**
Graph Cuts, P^n Model	0.67	0.77	0.28	1320

(b) F-Measure and Running Time on Tougher Dataset-I (95 targets, 5 slices)

Method	Level Set Tracking	HGraph3D	HGraph Propagate	P^n Graph Cuts
Median	0.87	**0.91**	0.89	**0.91**
RunTime(sec.)	378	210	**61**	950

(c) F-Measure and Running Time on Easier Dataset-II (30 targets, 10 slices)

Method	Frame1	Frame2	Frame3	Frame4	Frame5
Level Set Tracking	0.82	0.78	0.75	0.72	0.69
BiGraph3D	0.81	0.70	0.64	0.60	0.58
HGraph3D (Proposed)	**0.88**	**0.86**	0.83	0.79	0.77
HGraph Propagate (Proposed)	**0.88**	**0.86**	**0.85**	0.83	**0.80**
Graph Cuts, P^n Model	-	0.81	**0.85**	0.84	**0.80**

(d) Rand Indices on Tougher Dataset-I (95 targets, 5 slices)

Fig. 3. Validation of the Proposed Tracing Framework

and thus performs a one shot optimization. On the other hand, HGraph Propagate attempts to propagate contours in a slice by slice manner with segmentation of one slice being the prior for subsequent slice. Figure 3(a) illustrates the result of tracing all structures over the first few frames of the dataset. In order to place the proposed algorithm in context with existing state of the art techniques, we compare performance with Graph Cuts based P^n model [6], Level Sets using the Chan-Vese model, and bi-graph diffusion, see Figure 3. *The striking aspect of experiments is the running time of algorithms. For instance, HGraph Propagate runs in 34 seconds on Dataset-I without compromising accuracy, in comparison to graph cuts which takes 1320 seconds. This is a speed up of almost 35×.* Table 3(b),3(c) reports statistics on F-Measures against the ground truth over 475 (Dataset-I) and 300 (Dataset-II) contours

respectively on two different datasets with corresponding run times. Dataset-I is much more challenging due to appearance variability and larger number of labels. Similarly, Table 3(d) also reports the average Rand index of Dataset-I over all structures in every frame. In conclusion, this paper presented a simple tracing technique that easily scales to hundreds of labels. Experimental results on electron micrographs and comparisons to state of the art illustrate the method's applicability. Future work includes stable auto seeding using Minimum Description Length, deployment on larger distributed computing infrastructures, and active learning based interactive tracing.

References

1. Anderson, J., Mohammed, S., Grimm, B., Jones, B., Koshevoy, P., Tasdizen, T., Whitaker, R., Marc, R.: The viking viewer for connectomics: scalable multi-user annotation and summarization of large volume data sets. Journal of Microscopy 241(1), 13–28 (2011)
2. Boyd, S., Vandenberghe, L.: Convex optimization. Cambridge Univ. Press (2004)
3. Cormen, T., Leiserson, C., Rivest, R., Stein, C.: Introduction to Algorithms. The MIT Press (2001)
4. Dollar, P., Tu, Z., Belongie, S.: Supervised learning of edges and object boundaries. In: CVPR, pp. 1964–1971. IEEE Computer Society (2006)
5. Huang, Y., Liu, Q., Metaxas, D.: Video object segmentation by hypergraph cut. In: CVPR, pp. 1738–1745. IEEE (2009)
6. Jagadeesh, V., Vu, N., Manjunath, B.S.: Multiple Structure Tracing in 3D Electron Micrographs. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part I. LNCS, vol. 6891, pp. 613–620. Springer, Heidelberg (2011), http://dx.doi.org/10.1007/978-3-642-23623-5_77
7. Jurrus, E., Hardy, M., Tasdizen, T., Fletcher, P., Koshevoy, P., Chien, C., Denk, W., Whitaker, R.: Axon tracking in serial block-face scanning electron microscopy. Medical Image Analysis 13(1), 180–188 (2009)
8. Kaynig, V., Fuchs, T.J., Buhmann, J.M.: Geometrical Consistent 3D Tracing of Neuronal Processes in ssTEM Data. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part II. LNCS, vol. 6362, pp. 209–216. Springer, Heidelberg (2010), http://dx.doi.org/10.1007/978-3-642-15745-5_26
9. Ojala, T., Pietikäinen, M., Mäenpää, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. PAMI 24(7), 971–987 (2002)
10. Rabiner, L.: A tutorial on hidden markov models and selected applications in speech recognition. Proceedings of the IEEE 77(2), 257–286 (1989)
11. Shih, M.C., Rose, K.: Hidden markov models for tracking neuronal structure contours in electron micrograph stacks. In: ISBI, pp. 1377–1380. IEEE (2012)
12. Vazquez-Reina, A., Gelbart, M., Huang, D., Lichtman, J., Miller, E., Pfister, H.: Segmentation fusion for connectomics. In: ICCV, pp. 177–184. IEEE (2011)
13. Zhou, D., Bousquet, O., Lal, T., Weston, J., Schölkopf, B.: Learning with local and global consistency. In: NIPS, vol. 16, pp. 321–328 (2004)
14. Zhou, D., Huang, J., Scholkopf, B.: Learning with hypergraphs: Clustering, classification, and embedding. In: NIPS, vol. 19, p. 1601 (2007)

A Diffusion Model for Detecting and Classifying Vesicle Fusion and Undocking Events

Lorenz Berger[1], Majid Mirmehdi[2], Sam Reed[3], and Jeremy Tavaré[3]

[1] Department of Computer Science, University of Oxford, UK
[2] Department of Computer Science, University of Bristol, UK
[3] Department of Biochemistry, University of Bristol, UK

Abstract. Fluorescently-tagged proteins located on vesicles can fuse with the surface membrane (visualised as a 'puff') or undock and return back into the bulk of the cell. Detection and quantitative measurement of these events from time-lapse videos has proven difficult. We propose a novel approach to detect fusion and undocking events by first searching for docked vesicles that 'disappear' from the field of view, and then using a diffusion model to classify them as either fusion or undocking events. We can also use the same searching method to identify docking events. We present comparative results against existing algorithms.

1 Introduction

The movement of numerous proteins between the various sub-compartments of a cell is critical in the biological function of a cell. Defects in protein movement can lead to disease, e.g. ineffective movement of a protein called GLUT4 from small intracellular vesicles towards the surface membrane of a fat and muscle cell, and their consequent fusion with that membrane, leads to insulin resistance and type 2 diabetes [1]. Understanding these processes at a molecular level is, therefore, critical to understanding cellular behavior in normal and diseased states.

The docking and fusion of intracellular vesicles with the surface membrane of cells can be visualised using Total Internal Reflection Fluorescence Microscopy (TIRFM). A snapshot of the distribution of vesicles labeled with a fluorescently-tagged GLUT4 in a single fat cell is shown in Fig. 1(left) while Fig. 1(right) shows some of the key events in GLUT4 movement to the surface membrane. These key events in vesicle movement to the surface membrane can be described as follows. First insulin, which is required for the vesicles to fuse with the membrane, signals for the vesicles to move towards the surface membrane. Then some of the vesicles make it to the edge of the cell and 'dock' with the cell membrane. This docking event corresponds to vesicles suddenly halting and vibrating in the same place for a few seconds. After docking for some time, some of the vesicles then fuse at the cell membrane. This can be seen as a 'puff' (see row A in Fig. 2). Other vesicles dock for a few seconds and then undock and leave the vicinity of the membrane, returning back into the main bulk cell (i.e. go out of view; row B in Fig. 2) or move off to a different part of the membrane.

Extracting information and quantifying vesicle dynamics in TIRFM videos has proven difficult and is a major barrier to understanding their molecular basis. It is impractical to manually mark the video data as it would be far too time-consuming and
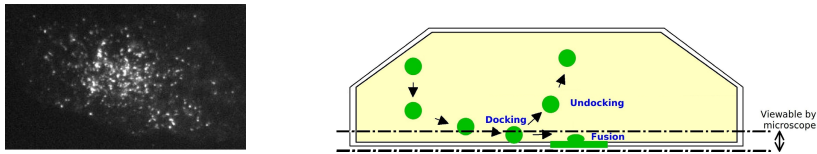
Fig. 1. (left) A fat cell's vesicles (bright dots) tagged with a green fluorescent protein and imaged using TIRFM. (right) Key events of insulin-stimulated GLUT4 translocation (see text for details).

error-prone, making an automatic approach necessary. Developing a robust quantitative vesicle analysis method will be important in many areas of cell biology where vesicle fusion and undocking with the surface membrane occurs, e.g. events such as neurosecretory vesicle fusion [2], which is relevant to Alzheimer's disease and schizophrenia. Current methods that are popularly and extensively used for biological research first segment individual vesicles by analyzing the surrounding gray level distribution, and then use thresholds on the pixel intensity to distinguish between fusion and undocking events, e.g. Bai et al.[3] and Huang et al. [4]. In Vallotton et al. [5], a matched filtering approach is used to identify events that highly correlate against a standard fusion event through space and time. However, this approach is not only computationally expensive, but tends to miss many events due to the high variability in duration, size and noise levels of fusion events. In Mele et al. [6], videos are represented as a 3D image space, where using a threshold for noise, patches of interest are differentiated from noise according to absolute difference in pixel intensity between frames. These patches of interest are then further analysed using a 3D extension of the Maximally Stable Extremal Regions algorithm to detect high intensity and highly variable regions which would correspond to potential fusion events. A set of descriptors including intensity differences, fusion spot size, and degree of fit to a diffusion model are collected for each candidate event, and are then compared against pre-identified fusion events using PCA. *This method however does not consider undocking events.*

2   Proposed Method

We first outline a search algorithm for detecting vesicles that suddenly disappear due to fusion or undocking - we shall call these candidate events. By modifying this algorithm we are also able to detect vesicles that suddenly appear, and to the best of our knowledge this is the first approach that has been proposed to detect docking events. To classify the candidate events, we build on an idea taken from [6] which is to use a mathematical diffusion model to extract features for classification. In [6] an analytical solution to the diffusion equation is derived to explain how the intensity of a vesicle should change over time during fusion (recall as mentioned earlier, [6] cannot handle undocking, and docking, events). We develop this idea by considering a 2D diffusion model which considers the whole vesicle and its surrounding area, and we also introduce a source and sink term to model both fusion and undocking events explicitly. We then fit these fusion and undocking models to the candidate event and use the goodness of the fit

to classify the event as a fusion or an undocking event. We test our algorithm on real TIRFM data and demonstrate its performance for a range of videos.

Detecting Fusion and Undocking Events - Sophisticated algorithms to track the behaviour of vesicles already exist, e.g. [7]. Here, the vesicles of interest are only those that have docked to the membrane and then go on to either fuse or undock, and thus, there is no need to track individual vesicles - this same phenomenon was also looked at in [8] using a patch-based method. Our main assumption is that a vesicle remains stationary for around N frames before it fuses or undocks. Hence, we search for vesicles that have been visible in the same position throughout the past N_p frames but then cannot be found again in this position during N_f future frames. This corresponds to the vesicle first being docked (visible) and then having either undocked or fused (not visible). We also allow for brief periods of vesicle disappearance. This allows us to detect vesicles that shortly go out of view due to noise and TIRFM artifacts but still follow the overall pattern of a disappearing vesicle. As preprocessing, we reduced noise in our videos using a standard Gaussian filter ($\sigma = 1.5$), as in [6]. For further efficiency, we also applied a low threshold to eliminate areas with very low intensity and no activity. We use local maxima to identify the individual vesicles in each frame which gives better results than using an adaptive thresholding approach as performed in [7].
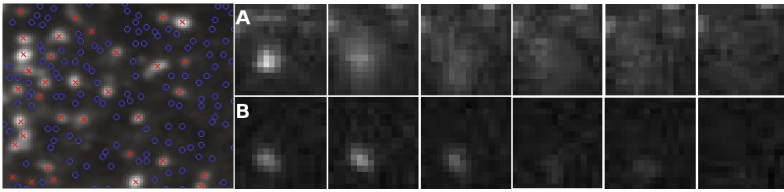
Fig. 2. (Left) Local maxima (red crosses) and local minima (blue circles). (Right) An image sequence showing a very prominent fusion event (row A), and an undocking event (row B).

Let $\mathbf{V}$ denote a 2D matrix that contains the positions of all local pixel maxima (see Fig. 2) in a 5×5 neighbourhood for every frame in the video, such that v_{ij} is the position of the jth maxima (vesicle) in the ith frame, and $\mathbf{v}_{k:} = [v_{k1}, v_{k2}, ..., v_{kj}, ...]$ denote the set of all the local maxima found in the kth frame. Similarly let $\mathbf{W}$ denote a 2D matrix that contains the positions of all the pixels that have a local minima (see Fig. 2) in a 3×3 neighbourhood in each frame such that w_{ij} is the position of the jth minimum in the ith frame. We can then calculate a score a_{ij} for each vesicle v_{ij} which reflects how closely v_{ij} follows the pattern of being stationary and visible for the past N_p frames and then being out of view for the future N_f frames:

$$a_{ij} = \sum_{k=i-N_p}^{k=i-1} h(\mathbf{v}_{k:}, v_{ij}) + \sum_{k=i+1}^{k=i+N_f} g(\mathbf{w}_{k:}, \mathbf{v}_{k:}, v_{ij}). \tag{1}$$

Here $h(\mathbf{v}_{k:}, v_{ij})$ is a function which searches $\mathbf{v}_{k:}$ for a local maxima that is positioned within r pixels of v_{ij}. More specifically,

$$h(\mathbf{v}_{k:}, v_{ij}) = \begin{cases} 1 & \text{if } \exists\, v \in \mathbf{v}_{k:} \text{ s.t } |v - v_{ij}| \le r, \\ 0 & \text{otherwise,} \end{cases} \tag{2}$$

where r is the radius of the vesicle. The function $g(\mathbf{w}_{k:}, \mathbf{v}_{k:}, v_{ij})$ searches $\mathbf{w}_{k:}$ for a local minima that is positioned within r pixels of v_{ij} and also checks that there are no local maxima $v \in \mathbf{v}_{k:}$ which are close to v_{ij}, i.e.

$$g(\mathbf{w}_{k:}, \mathbf{v}_{k:}, v_{ij}) = \begin{cases} 1 & \text{if } \exists\, w \in \mathbf{w}_{k:} \text{ s.t } |w - v_{ij}| \le r, \\ & \quad \wedge\ \forall\, v \in \mathbf{v}_{k:} \text{ s.t } |v - v_{ij}| > r, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

To decide whether a vesicle should qualify as having gone missing we set a threshold such that if $a_{ij} \ge (N_p + N_f)C$, then a missing vesicle is detected at v_{ij}, where $C \in [0, 1]$ is a threshold representing the amount of 'brief disappearance' that is allowed.

Detecting Docking Events - To detect vesicles that are first not visible but then suddenly appear and stay stationary we can simply rearrange Eq. (1) such that

$$b_{ij} = \sum_{k=i-N_p}^{k=i-1} g(\mathbf{w}_{k:}, \mathbf{v}_{k:}, v_{ij}) + \sum_{k=i+1}^{k=i+N_f} h(\mathbf{v}_{k:}, v_{ij}). \tag{4}$$

To decide whether a vesicle should qualify as having docked we set a threshold such that if $b_{ij} \ge (N_p + N_f)C$, then a docking vesicle is detected at v_{ij}.

3   A Computational Model for Fusion and Undocking Events

TIRFM helps create an evanescent field which illuminates and excites fluorophores in a region $\approx 100nm$ below the interface. The evanescent field's intensity decreases exponentially with the distance perpendicular to the interface which directly relates to an exponential decrease in the fluorescence [9]. For a vesicle that goes from being docked at the membrane to fusing and diffusing into the membrane, the total number of fluorophores does not change, however, as the fluorophores diffuse into the membrane they are now collectively closer to the interface which can result in a slight total intensity increase in the video. In the fusion model, this is represented using a diffusive and a source term centred at the vesicle. During an undocking event, there is no diffusion since the vesicle just undocks and returns into the cell. This is modelled using just a sink term centred at the vesicle which is able to explain the sudden decrease in intensity.

Definition of the Models - For each of the previously detected candidate events v_{ij} a sequence of subregions centred at the event spatially and temporally is taken from the video. A sequence length of $N_p/2 + N_f/2$ guarantees that the actual event happens during these frames. The size of the subregion should include the whole vesicle and

some room for a fusion (puff) to happen - we take this to be $4r \times 4r$. Let $\mathbf{I}_k(x, y)$ be the pixel intensity at position (x, y) within the subregion, during the kth frame, with $x \in [-2r, 2r]$, $y \in [-2r, 2r]$ and $k \in [i - N_p/2, i + N_f/2]$. Let $\mathbf{M}_k$ be the circular mask of radius r, centred at the candidate vesicle position $v_{ij} = (v_x, v_y)$ such that

$$\mathbf{M}_k(x, y) = \begin{cases} \mathbf{I}_k(x, y) & \text{if } (x - v_x)^2 + (y - v_y)^2 \leq r^2 \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

This will be the mask on which the source or sink will be able to act. We can then introduce the model for a fusion event as

$$\frac{\partial \mathbf{I}}{\partial t} = D_F \left(\frac{\partial^2 \mathbf{I}}{\partial x^2} + \frac{\partial^2 \mathbf{I}}{\partial y^2} \right) + S_F \mathbf{M}_k, \tag{6}$$

where $S_F \in [0, \infty)$ is the magnitude of the source and $D_F \in [0, \infty)$ is the diffusion coefficient, which model the amount of increase and diffusion of fluorescence respectively. The model for an undocking event is given by

$$\frac{\partial \mathbf{I}}{\partial t} = S_U \mathbf{M}_k, \tag{7}$$

where $S_U \in (-\infty, 0]$ is the magnitude of the sink. As there is no puff, i.e. no diffusion in an undocking event, then $D_U = 0$, causing the diffusion term to disappear.

Using the Models for Classification - We can use Eqs. (6) and (7) to explain how a candidate event evolves temporally and whether it can be classified as a fusion event, an undocking event, or neither. We pick a frame as initial conditions to (6) and (7), evolving the system by one time step and then comparing the predicted result with the actual next frame. The total intensity difference between the predicted frame and the actual frame of each model is then turned into a likelihood ratio which is used for classification. Let us define a function which evolves the fusion model in (6) by one time step and solves for the new intensity distribution $\hat{\mathbf{I}}_{k+1}^F$ of the subregion

$$\hat{\mathbf{I}}_{k+1}^F = \mathbf{F}(\mathbf{I}_k, D_F, S_F). \tag{8}$$

The fusion model in (6) is solved numerically using the Crank-Nicolson method, with homogeneous Neumann boundary conditions and initial conditions $\mathbf{I}_k$. We then optimize (8) to find the optimal D_F and S_F values which let the fusion model best predict the next frame. This optimization step is implemented using the well known Nelder-Mead method. The absolute difference ξ_{Fk} over the whole subregion between the predicted frame $\hat{\mathbf{I}}_{k+1}^F = \mathbf{F}(\mathbf{I}_k, D_F, S_F)$ and the actual frame $\mathbf{I}_{k+1}$ is then given by $\xi_{Fk} = \min_{D_F, S_F} \int_\Omega |\hat{\mathbf{I}}_{k+1}^F - \mathbf{I}_{k+1}| \, d\Omega$, where Ω is the area of the subregion and $k = [i - N_p/2, ..., i + N_f/2]$. A similar calculation can be done for the undocking model using $\xi_{Uk} = \min_{S_U} \int_\Omega |\hat{\mathbf{I}}_{k+1}^U - \mathbf{I}_{k+1}| \, d\Omega$, where $\hat{\mathbf{I}}_{k+1}^U = \mathbf{U}(\mathbf{I}_k, S_U)$ is the function that evolves the undocking model (7) by one time step and solves for a new intensity distribution of the subregion. We can then define a likelihood ratio $\lambda_k = \frac{\xi_{Uk}}{\xi_{Fk}}$ to see which model is better at predicting the next frame. When $\lambda_k > 1$, then the fusion model is better at predicting the next frame $\mathbf{I}_{k+1}$,

and if $\lambda_k < 1$, then the undocking model is better. To finally decide whether a fusion event has happened, we use a simple threshold on λ to determine the class, i.e. *Fusion* if $(\max(\lambda) \geq \alpha)$, *Undocking* if $(\min(\lambda) < \gamma \wedge \max(\lambda) < \alpha)$, or *Noise* otherwise. The thresholds can be determined by inspecting the results of a few known events.

4   Results

All the results presented here are evaluated against groundtruth generated by a cell biology expert working in the field of TIRF microscopy and vesicle trafficking. Our data set comprises of four videos[1]. The first part of the proposed method, which searches for 'disappearing' vesicles, achieves an average detection rate of 87.5%. This might not seem as high as expected but is due to the inherent difficulty of the problem. The fusions can be extremely varied in their nature and are often, even for biology experts, difficult to spot and classify. The low spatial and temporal resolution and low SNR in the videos, due to microscopy limitations, also add to the difficulty of the problem. Table 1 shows the number of fusion and undocking events in the groundtruth for each video, the total number of candidate events detected by the proposed method, and the number of candidate events detected by the proposed method corresponding to true events in the groundtruth. In these experiments, we set $N_p = N_f = 20$ and $C = 0.6$. Other advantages of the proposed method are that it is easy to implement, it is fast, and it does not rely on any intensity thresholds, just minima and maxima. This part of the proposed method, i.e. the detection of candidate events alone can be very useful for biologists who analyze such videos. What previously took a full day's work to analyze one video manually can now be achieved in just a few minutes by finding all candidate events in a video (about 0.25 fps) and then manually classifying the events as fusion or undocking events. To automatically classify the events, a threshold on the likelihood ratio produced by the two models is used. The large peak in Fig. 3(left) demonstrates the algorithm's ability to produce a clear signal even for cases where a fusion looks fairly similar to an undocking event. This large peak can then easily be classified as a fusion event.

Table 1. Precision and Recall results for detecting candidate events

	Fusion events groundtruth	Undocking events groundtruth	Candidate events detected	Correct candidate events	Recall	Precision	Time taken seconds (s)
Movie1	16	7	24	19	82.6%	79.2%	42 s
Movie2	10	3	13	13	100%	100%	30 s
Movie3	15	12	35	22	81.5%	62.9%	43 s
Movie4	9	12	23	18	85.7%	78.3%	31 s
Average					87.5%	80.1%	

In Fig. 3(right), an example of an undocking event is shown, where the large trough in the likelihood ratio can also be easily classified as an undocking event. The vesicle in

---

[1] Note in other works, e.g. [3] and [4], only one long movie was tested.
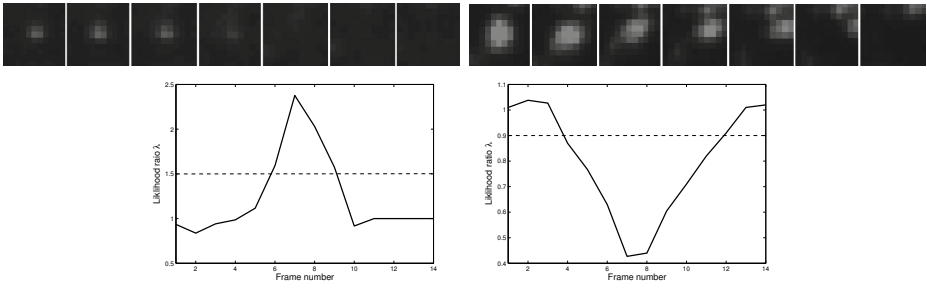
Fig. 3. (Top left) frames 4-11 taken from a fusion event in Movie3.avi, (bottom left) the likelihood ratio between the fusion model and the undocking model for different frames during the fusion sequence, and a threshold of $\alpha = 1.5$. (Top right) an undocking event and its likelihood ratio (bottom right) with a threshold of $\gamma = 0.9$.

the image sequence is stationary and then moves away (undocks) instead of returning straight back into the cell which is more common for an undocking event (see Fig. 2(B)). This is where other methods, such as [3] that use intensity thresholds on features including the maximum intensity increase in the annular area of the vesicle, would fail and incorrectly classify the event as fusion due to the increase in intensity in the annulus, which of course is caused by the vesicle movement, and not by a fusion.

Table 2. Results for the automatic classification of detected events

Fusion	Accuracy	Recall	Precision	False +ve	Undocking	Accuracy	Recall	Precision	False +ve
Movie1	91.7%	100.0%	85.7%	16.7%	Movie1	91.7%	85.7%	85.7%	5.9%
Movie2	76.9%	80.0%	88.9%	33.3%	Movie2	76.9%	33.3%	50.0%	10.0%
Movie3	80.0%	80.0%	61.5%	20.0%	Movie3	54.3%	60.0%	37.5%	43.5%
Movie4	91.3%	83.3%	83.3%	5.9%	Movie4	78.3%	75.0%	81.8%	18.2%
Average	85.0%	85.8%	79.9%	19.0%			75.3%	63.5%	63.8%	19.4%

Table 2 shows the results for the automatic classification of events detected during the searching stage with $\alpha = 1.5$ and $\gamma = 0.9$. It also shows that the diffusion model is able to correctly classify the majority of events and is consistent in detecting fusion events across videos, which are of prime interest to researchers. The complete analysis for a typical set of 200 frames of size 160x160 on a standard 2GHz processor took around 1-2 minutes when implemented in MATLAB. Results for detecting docking events have not been presented here because obtaining the groundtruth for them is extremely cumbersome. Fig. 4 shows a comparison of our method against Bai et al. [3] and Huang et al. [4] using ROC plots. The parameters in algorithms [3] and [4] were optimised for our videos to give their best possible results. Our proposed method outperforms [3] and [4], since it avoids the use of pixel intensity thresholds for classification. This makes it more robust to different videos with different quality and image properties as well as busy regions where closely neighbouring vesicles can interfere with events that are being analyzed. Fig. 4 (right) also shows a basic sensitivity analysis on the classification threshold α which has been performed over all videos. This analysis only looks at the
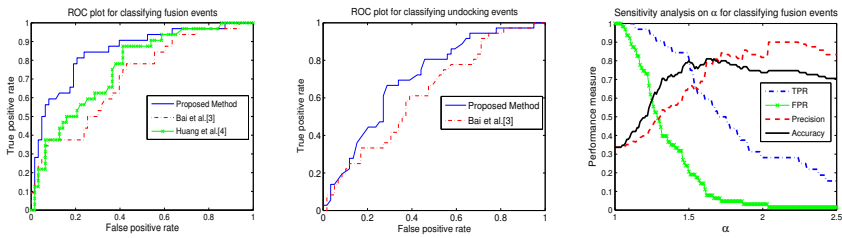
Fig. 4. ROC plots for comparative evaluation of fusion (left) and undocking (middle) events, and a sensitivity analysis (right) of α on the classification of fusion events

classification of fusion events which are of main interest. Choosing α between 1.3 and 1.8 seems to optimize most of the performance measures.

5 Conclusion

Quantitatively measuring the rate of fusions and undocking vesicles is a common problem in cell biology and crucial for making progress in researching the biological function of cells. We proposed a simple, fast and easy to implement search algorithm to find disappearing vesicles. This searching algorithm can also be reformulated to detect docking events. To automatically classify the disappearing events, we proposed a novel computational diffusion model for both fusion and undocking events.

References

1. Leney, S., Tavaré, J.: The molecular basis of insulin-stimulated glucose uptake: signalling, trafficking and potential drug targets. Journal of Endocrinology 203(1), 1–18 (2009)
2. Calì, C., Marchaland, J., Regazzi, R., Bezzi, P.: SDF 1-alpha (CXCL12) triggers glutamate exocytosis from astrocytes on a millisecond time scale: imaging analysis at the single-vesicle level with TIRF microscopy. Journal of Neuroimmunology 198(1-2), 82–91 (2008)
3. Bai, L., Wang, Y., Fan, J., Chen, Y., Ji, W., Qu, A., Xu, P., James, D., Xu, T.: Dissecting multiple steps of GLUT4 trafficking and identifying the sites of insulin action. Cell Metabolism 5(1), 47–57 (2007)
4. Huang, S., Lifshitz, L., Jones, C., Bellve, K., Standley, C., Fonseca, S., Corvera, S., Fogarty, K., Czech, M.: Insulin stimulates membrane fusion and GLUT4 accumulation in clathrin coats on adipocyte plasma membranes. Molecular and Cellular Biology 27(9), 3456–3469 (2007)
5. Vallotton, P., James, D., Hughes, W.: Towards fully automated identification of vesicle-membrane fusion events in TIRFM. In: AIP Proceedings, vol. 952, pp. 3–10 (2007)
6. Mele, K., Coster, A., Burchfield, J., Lopez, J., James, D., Hughes, W., Vallotton, P.: Automatic identification of fusion events in TIRF microscopy image sequences. In: ICCV Workshops Proceedings, pp. 578–584. IEEE (2009)
7. Singh, S., Bhaskar, H., Tavare, J., Welsh, G.: Multiple Particle Tracking for Live Cell Imaging with Green Fluorescent Protein (GFP) Tagged Videos. In: Singh, S., Singh, M., Apte, C., Perner, P. (eds.) ICAPR 2005. LNCS, vol. 3687, pp. 792–803. Springer, Heidelberg (2005)
8. Boulanger, J., Gidon, A., Kervran, C., Salamero, J.: A patch-based method for repetitive and transient event detection in fluorescence imaging. PLoS One 5(10), e13190 (2010)
9. Oheim, M., Schapper, F.: Non-linear evanescent-field imaging. Journal of Physics D: Applied Physics 38, R185 (2005)

Efficient Scanning for EM Based
Target Localization

Raphael Sznitman, Aurelien Lucchi, Natasa Pjescic-Emedji,
Graham Knott, and Pascal Fua

Computer Vision Lab,
École Polytechnique Fédérale de Lausanne, Switzerland
`firstname.lastname@epfl.ch`

Abstract. For biologists studying the morphology of cells, Electron Microscopy (EM) is the method of choice with its nm resolution. However, the time necessary to acquire EM image series is long and often limits both the number and size of samples imaged. This paper presents a strategy for fast imaging and automated selection of regions of interest that significantly accelerates this process. In particular, this strategy involves scanning a tissue sample once, finding regions of interest in which target structures might be located, scanning these regions once again, and iterating the process until only relevant regions of the block face have been scanned repeatedly. For mitochondria and synapses, this approach is shown to produce near equal localization results to current state-of-the art techniques, and does so in almost a tenth of the time.

1 Introduction

Focused Ion Beam Scanning Electron Microscopes (FIB-SEM) and their ability to image with isotropic resolution of up to $4nm$ per pixel are becoming invaluable tools in studying cell ultrastructure and model organelles, such as mitochondria, synapses, and vesicles. Acquiring images such as those depicted in Fig. 1 involves repeatedly milling a few nm from the surface of a tissue block using a focused gallium ion beam, scanning each line of a rectangular region of the block face several times, averaging the results, and milling again.

The resulting images have already yielded many new insights in the structure and functioning of cells [1, 2], but the acquisition process is desperately slow. For example, imaging the $10 \times 10 \times 10 \mu m$ tissue block of Fig. 1 at full resolution took approximately 50 hours. Such lengthy processing times are limiting because neuroscientists now require larger volumes to enable multiple cells, and even entire tissue samples, to be analyzed, which would currently be prohibitively slow. Furthermore, because thermal changes can cause the block face to drift and produce misaligned image series, considerable precision is needed to maintain consistent imaging. This is difficult to achieve over extremely long periods and limits the size of the images that can be captured.

In short, a pressing need exists to reduce scanning time without compromising the usefulness of the resulting images. Some research has already gone into
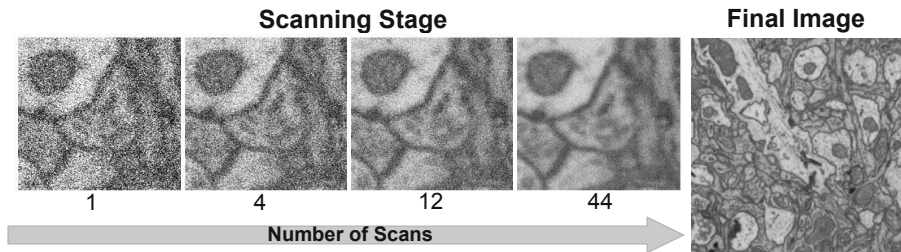
Fig. 1. FIB-SEM Scanning Process. The microscope repeatedly scans the surface of a sample block until a clear image is produced.

achieving such a result. For example, synthetic and real data have been used to show that Sparse Sampling techniques could potentially help [3]. However, the current generation of microscopes simply cannot perform the required random sampling because imaging extremely small and random locations on the block face would be incredibly time consuming.

In contrast, this paper uses a combination of real and synthetic data to demonstrate an approach to achieving much faster scanning using existing technology [4] when the images are intended for automated detection, counting, and modeling of organelles. The strategy involves scanning each image line of the block surface *once* instead of several times, finding rectangular regions of interest in which target structures might be located, scanning these regions once again, and iterating the process until only relevant regions of the block face have been scanned repeatedly. This process will be shown to result in much reduced imaging time at almost no performance loss.

The remainder of this paper is organized as follows. The paper begins by specifying our approach in Sec. 2, and provides a problem formulation in Sec. 2.1 and an algorithm description in Sec. 2.2. In Sec. 3, the method is validated experimentally. The paper concludes with final remarks in Sec. 4.

2   Sequential Region Cascades

The following observations are used as the starting point:

1. If the goal is to count or model organelles that occupy only a small fraction of a tissue block, precisely imaging the entire tissue block is a waste of time. This concept has long been known and exploited by Computer Vision researchers [5], [6] but, to the best of our knowledge, has been neglected by microscopists.
2. Once an organelle has been found in one slice, it will be seen with very high probability at similar locations in subsequent slices, *e.i.* organelles are 3D structures. Similarly, a slice must exist in which an organelle first appears, as well as a final one.

3. Current microscopes are good at quickly scanning rectangular regions, and new scanning engines that will make this process entirely programmable on a slice-by-slice basis will soon be commercially available [4]. In contrast, microscopes that can quickly and randomly access specific image locations are likely to remain beyond the state-of-the-art for the foreseeable future, thus precluding the use of standard Sparse Sampling techniques [3].
4. The microscopes are optimized to scan lines at one particular speed. Total scanning time, and image quality, are controlled by the number of times each line of the rectangular regions of interest is scanned. Typically, the final result is the average of these scans.

Therefore, we propose our Sequential Region Cascades (SRC) approach for accelerating the scanning process by building a cascade of region classifiers and sequentially evaluating regions that appear likely to contain a particular intracellular structure. While most cascade systems increase classification accuracy over levels, given the technical constraints discussed above, classification is instead performed on images acquired with varying scan counts. To take advantage of what was found in one slice and accelerate the scanning in the next, target locations are directly imaged using locations from previous slices with the largest possible numbers of scans, and new targets are searched for in the remainder of the slice using the cascade.

In the remainder of this section, the SRC approach is first formalized, and then its implementation is discussed. Section 3 shows that it allows counting of both mitochondria and synapses with much reduced acquisition times and at almost no loss in accuracy.

2.1 Problem Formulation

We formalize our problem as follows. Let the volume to image be denoted by $\mathcal{V} = \{S^1, \ldots, S^T\}$, where S^t corresponds to a slice of the volume. When using a scanning EM microscope, we consider two sets of parameters when acquiring images. First, we define a rectangular region to scan, $R = (r_1, r_2)$, where r_1 and r_2 are the upper left and bottom right pixel coordinates, respectively. Second, the *scan count* is defined as the number of times the electron beam images one pixel and denote this value as $C = \{1, \ldots, C_{\max}\}$.

Given these two parameters, (R, C), the process of acquiring an image by scanning a region of a slice can be described by the function f, where $f : S \times R \times C \mapsto I$. That is, evaluating the function $f(S, R, C)$ provides an image I_C^R of size R and corresponds to the average of C independently scanned samples. Typically, the time cost associated with evaluating $f(S, R, C)$ is $C \times area(R)$.

As in [7–9], we may train a classifier to verify if a pixel in I_C^R belongs to a target organelle. As can be seen in Fig. 1, images acquired using different scan counts exhibit different statistics and we may train different classifiers for each. We therefore take a family of classifier to be $\mathcal{H} = \{h_1, \ldots, h_{C_{\max}}\}$, where $h_c : I_C^R \mapsto \{0, 1\}^R$. Note that these classifiers return binary images.

Algorithm 1. Sequential Region Cascades (SRC)

1: $\mathcal{P} \leftarrow$ empty queue, $\mathcal{Q} \leftarrow$ empty queue.
2: **for** $t = 1, \ldots, T$ **do**
3:     Push$(([0,0], [M,N]), 1)$ into $\mathcal{Q}$
4:     RemoveOverlaps$(\mathcal{Q}, \mathcal{P})$
5:     **while** $|\mathcal{Q}| > 0$ **do**
6:         $[R, C] \leftarrow$ Top$(\mathcal{Q})$
7:         $I_C^R \leftarrow f(S^t, R, C)$
8:         $\mathcal{R} \leftarrow$ ExtractRegions$(h_C(I_C^R))$
9:         **if** $C_{\max} == C$ **then**
10:             Push$(\mathcal{R}, C)$ into $\mathcal{P}$
11:         **else**
12:             Push$(\mathcal{R}, C+1)$ into $\mathcal{Q}$
13:         **end if**
14:     **end while**
15: **end for**

Finally, let P^t to be the set of pixels corresponding to the location of target structures in slice S^t. Our goal is to discover these sets $\{P^t\}_{t=1}^{t=T}$ for all slices as efficiently and as quickly as possible.

2.2 Algorithm and Implementation

An outline of the algorithm is shown in Alg. 1. To begin, the user provides the set of classifiers and the volume to image, $\mathcal{H} = \{h_1, \ldots, h_{C_{\max}}\}$ and $\mathcal{V} = \{S^1, \ldots, S^T\}$, respectively. The algorithm begins by forming two queues that will maintain tuples of regions and scan counts, *i.e.* (R, C). The first, $\mathcal{P}$, maintains a set of regions deemed the target structure on a given slice. The second, $\mathcal{Q}$, maintains an intermediary list of candidate regions that appear likely to contain target structures within them. Initially, both queues are empty.

For each new slice, we begin by pushing the entire observable domain as a candidate region using the smallest scan count. The following sequence of steps is then looped, which is called the *refining stage* and is depicted by Fig. 2 (lines 5 to 14): A candidate region and scan count index from the queue $\mathcal{Q}$ is retrieved. The associated image region, I_C^R, is then acquired and the corresponding classifier is evaluated by computing $h_c(I_C^R)$. At this point, the binary classification image is searched for disjoint sets of rectangular regions that indicate potential target locations. If newly extracted regions were acquired using the highest possible scan count, these are pushed into the target region queue, $\mathcal{P}$, otherwise, they are pushed into the candidate region queue, $\mathcal{Q}$. This process iterates until no candidate region remains.

In addition, before starting the *refining stage*, the overlapping region that coexists in both $\mathcal{Q}$ and $\mathcal{P}$ are removed. This effectively creates a new set of regions in $\mathcal{Q}$ that are disjoint of $\mathcal{P}$ and reduces the direct need for searching targets likely to have stayed in the same location. Doing so is one way of encoding
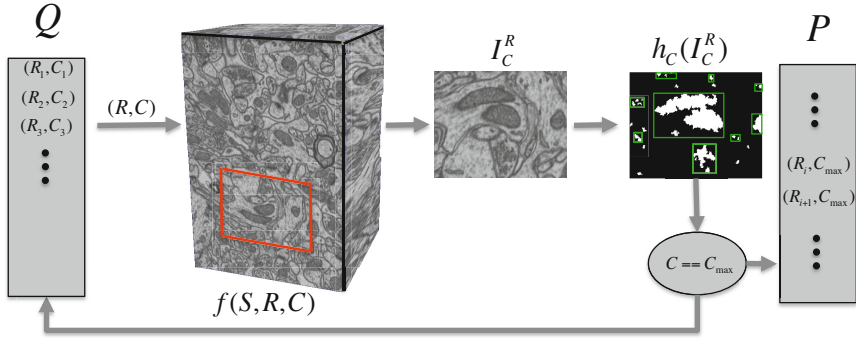
Fig. 2. SRC Refining Stage: For any slice, the queue $\mathcal{Q}$ contains a list of tuples (R, C) that specify a region and scan count with which the microscope should image. Once an image region acquired, I_C^R, a dedicated classifier assigns a binary label to each location in the image for the presence of targets. New tuples are then formed and inserted into $\mathcal{Q}$ or into the permanent queue $\mathcal{P}$.

3D information for location of targets in a volume. Obviously, during the first slice of the tissue block, this step is irrelevant because $\mathcal{P}$ is empty.

3   Experiments and Results

We used a Zeiss NVision40 FIB-SEM microscope to mill and scan a rodent brain sample of $10 \times 10 \times 1\mu m$, which produced 165, 1024×1536 images. We also collected a second stack of 377, 655×429 images. In both cases, each line was scanned $C_{\max} = 44$ times. This took 5 and 12 hours, respectively. We evaluated our algorithm for the tasks of localizing two types of organelle: mitochondria and synapses. Here, we show how the SRC strategy could be used to divide the scanning time by a factor of 10 to 15 depending on the target type.

Test Data: Given that the scanning engine of our NVision40 microscope was not designed for this, implementing our approach on it would be very difficult. However, this will soon change when newer scanning engines come to market [4]. To demonstrate our approach in the meantime, we therefore proceed as in [3] and synthesized the scans we would have gotten using values of $C < C_{\max}$ by appropriately degrading the higher-quality ones. To ensure realism we proceeded as follows.

Before scanning, and thus destroying, the whole first block, we collected six independent images $\{I_1, ..., I_6\}$ of the first slice using a single scanning pass for each and a single one, which we denote as $\hat{I}$, using 6 passes. From this, we first verified that for any pixel location u, the gray-level $I_i(u)$ in any of the 6 single-scan images is well approximated by a Gaussian of mean $\hat{I}(u)$ and standard deviation $\sigma_u = m\hat{I}(u) + b$, where m and b are linear regression parameters, as
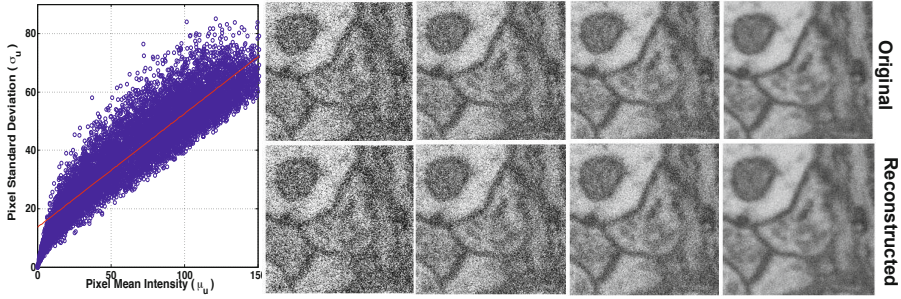
Fig. 3. Reconstructing small scan count images. (left) Relation between Average intensity and standard deviation for pixels acquired with low scan counts. (right) Top rows shows real images acquired under 1, 4, 12 and 44 scan counts, and the bottom shows the corresponding reconstructed images.

illustrated by Fig. 3 where we plot μ_u against σ_u for 10'000 randomly selected pixels. In other words, the gray level variance is directly proportional to the gray level value.

From this, using an image acquired with a large number of scans, we can simulate acquiring an image from a smaller number of scans. For example, a pixel with n scans can be reconstructed by sampling the Gaussian $\mathcal{G}(\hat{I}(u), m\hat{I}(u) + b)$ n times and averaging the samples. Fig. 3 shows the true (top row) and reconstructed (bottom row) images using this process with 1, 4, 12 and 44 scan counts.

Experimental Setup: We tested four scan count sequences—$Cs = \{12, 44\}$, $Cs = \{6, 12, 44\}$, $Cs = \{1, 6, 12, 44\}$, $Cs = \{1, 4, 6, 44\}$—for synapse and mitochondria detection purposes. Each classifiers h_c were built as in [7]. That is, we first extracted regularly spaced superpixels from which we computed both intensity histograms and steerable features. We then used 15 training images for each scan count C to train a different Support Vector Machine (SVM) classifier with an RBF kernel. In the $C = C_{\max}$ case, we used the original images and for all other $C < C_{\max}$, we used synthesized images obtained as discussed above.

Fig. 4 depicts the algorithm's behavior when attempting to locate mitochondria using the scan sequence $Cs = \{6, 12, 44\}$. From left to right, we show the complete set of regions evaluated with each scan count on the initial slice, S^1. In red, we show $C = 6$, in green $C = 12$ and in blue $C = 44$. The ground truth regions are also shown. Additionally, for $C = 12$ and 44, we also display what regions from the previous step were evaluated.

Evaluation: Recall that our goal is to image at full resolution all the regions containing target structures–mitochondria and synapses in our experiments–while spending as little time as possible scanning the block faces. In this context,
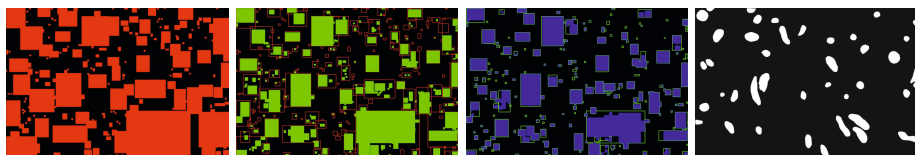
Fig. 4. SRC Examples. Images depict the algorithm's behaviour when attempting to locate mitochondria using the scan sequence $Cs = \{6, 12, 44\}$. From left to right, we show the complete set of regions evaluated with each scan count on the initial slice, S^1. In red, we show $C = 6$, in green $C = 12$ and in blue $C = 44$. The ground truth regions are also shown. Additionally, for $C = 12$ and 44, we also display what regions from the previous step were evaluated.

the proper measure of success is the True Positive Rate (TPR) as a function of scanning time, which is plotted in Fig 5. The false positive rate is less relevant as false positives only cause irrelevant parts of the block being scanned, which implies no loss of information but a time penalty that the increased scanning time already reflects.

For the purpose of this evaluation, we consider the time cost of one imaging strategy to be the sum of the number of times each individual pixel is scanned. For simplicity's sake, we normalize these numbers by the corresponding count when scanning the whole block at the maximum scan count. As a result, the times that appear in Fig 5 are numbers between 0 and 1.

From these results, we can see all the scan count sequences we tested provide a significant speed increase, mostly at a very small loss in TPR. By choosing the appropriate sequence we can establish more than five-fold speedups for an insignificant TPR loss. In practice, this means a neuro-scientist could examine and gather statistics for five times as many synapses in the same scanning time.

Note that the choice of which specific scan count sequence to use is not innocuous as it implies training different classifiers for each scan count value, some of which might be more appropriate than others. For example, the {1,4,6,44} sequence appears to outperform the {1,6,12,44} one, which may imply that additional research into optimizing these sequences might lead to further gains.
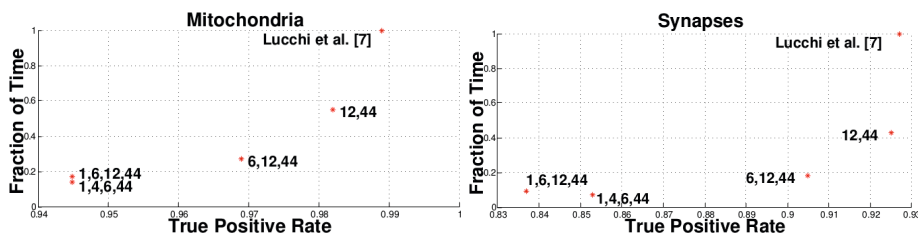
Fig. 5. Mean relative time and TPR for localizing Mitochondria and Synapses by tested methods. Each point indicates a specific scanning strategy. See text for details.

Also, the gain in time achieved by our approach is highly dependent on the type of organelle that must be found. In particular, targets that cove smaller surface areas, such as synapses (roughly 0.05% of the surface) allow higher time gains then for mitochondria (3 to 5% surface covered).

4   Conclusion

We presented an approach for speeding-up image acquisition when tasked with localizing specific structures in FIB-SEM imagery. It exploits the fact that low-quality images can be acquired faster than higher-quality ones and yet be sufficient for inference purposes. We have demonstrated greater than five-fold speed-ups at very little loss in accuracy in the context of mitochondria and synapse detection. Furthermore, the algorithm we propose is generic and applicable to many imaging modalities that allow trading quality for speed.

References

1. Knott, G., Marchman, H., Wall, D., Lich, B.: Serial Section Scanning Electron Microscopy of Adult Brain Tissue Using Focused Ion Beam Milling. Journal of Neuroscience 28(12), 2959–2964 (2008)
2. Kreshuk, A., Straehle, C., Sommer, C., Koethe, U., Knott, G., Hamprecht, F.: Automated segmentation of synapses in 3D EM data. In: 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, March 30-April 2, pp. 220–223 (2011)
3. Veeraraghavan, A., Genkin, A., Vitaladevuni, S., Scheffer, L., Xu, S., Hess, H., Fetter, R., Cantoni, M., Knott, G., Chklovskii, D.: Increasing depth resolution of electron microscopy of neural circuits using sparse tomographic reconstruction. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1767–1774 (June 2010)
4. Fibics, http://www.fibics.com/
5. Viola, P., Jones, M.: Robust Real-Time Face Detection 57(2), 137–154 (2004)
6. Sznitman, R., Jedynak, B.: Active Testing for Face Detection and Localization 32(10), 1914–1920 (2010)
7. Lucchi, A., Smith, K., Achanta, R., Lepetit, V., Fua, P.: A Fully Automated Approach to Segmentation of Irregularly Shaped Cellular Structures in EM Images. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part II. LNCS, vol. 6362, pp. 463–471. Springer, Heidelberg (2010)
8. Kaynig, V., Fuchs, T., Buhmann, J.: Neuron geometry extraction by perceptual grouping in sstem images. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 2902–2909 (June 2010)
9. Kumar, R., Vazquez-Reina, A., Pfister, H.: Radon-like features and their application to connectomics. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, pp. 186–193 (June 2010)

Automated Tuberculosis Diagnosis Using Fluorescence Images from a Mobile Microscope

Jeannette Chang[1], Pablo Arbeláez[1], Neil Switz[2], Clay Reber[2], Asa Tapley[2,3], J. Lucian Davis[3], Adithya Cattamanchi[3], Daniel Fletcher[2], and Jitendra Malik[1]

[1] UC Berkeley Department of Electrical Engineering and Computer Sciences
[2] UC Berkeley Department of Bioengineering
[3] UC San Francisco Medical School and San Francisco General Hospital

Abstract. In low-resource areas, the most common method of tuberculosis (TB) diagnosis is visual identification of rod-shaped TB bacilli in microscopic images of sputum smears. We present an algorithm for automated TB detection using images from digital microscopes such as CellScope [2], a novel, portable device capable of brightfield and fluorescence microscopy. Automated processing on such platforms could save lives by bringing healthcare to rural areas with limited access to laboratory-based diagnostics. Our algorithm applies morphological operations and template matching with a Gaussian kernel to identify candidate TB-objects. We characterize these objects using Hu moments, geometric and photometric features, and histograms of oriented gradients and then perform support vector machine classification. We test our algorithm on a large set of CellScope images (594 images corresponding to 290 patients) from sputum smears collected at clinics in Uganda. Our object-level classification performance is highly accurate, with Average Precision of $89.2\% \pm 2.1\%$. For slide-level classification, our algorithm performs at the level of human readers, demonstrating the potential for making a significant impact on global healthcare.

1 Introduction

Though tuberculosis (TB) receives relatively little attention in high-income countries, it remains the second leading cause of death from infectious disease worldwide (second only to HIV/AIDS) [10]. The majority of TB cases may be treated successfully with the appropriate course of antibiotics, but diagnosis remains a large obstacle to TB eradication. Presently, the most common method of diagnosing patients with TB is visually screening stained smears prepared from sputum. Technicians view the smears with microscopes, looking for rod-shaped objects (sometimes characterized by distinct beading or banding) that may be *Mycobacterium tuberculosis*, the bacteria responsible for TB disease. Apart from the costs of trained technicians, laboratory infrastructure, microscopes and other equipment, this process suffers from low recall rates, inefficiency, and inconsistency due to fatigue and inter-evaluator variability [9]. Hence, with the advent
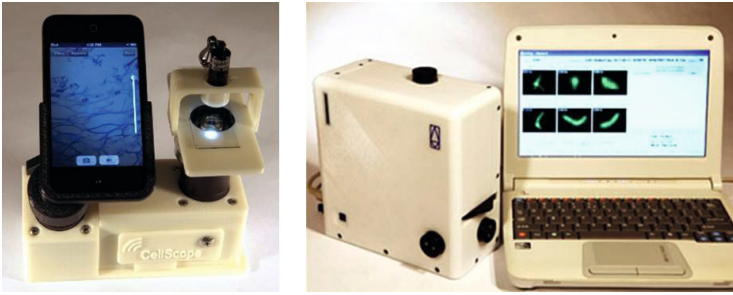
Fig. 1. Two versions of CellScope, a novel mobile microscope. Various uses include point-of-care diagnostics or transmission of images from rural areas to medical experts.

of low-cost digital microscopy, automated TB diagnosis presents a ready opportunity for the application of modern computer vision techniques to a real-world, high-impact problem.

We propose an algorithm for automated TB detection using images from digital microscopes such as CellScope [2] (Figure 1), a low-cost and portable alternative to standard laboratory-based microscopes. We present results from a large dataset of sputum smears collected under real-field conditions in Uganda. Our algorithm performs at the level of human readers when classifying slides, which opens exciting opportunities for deployment in large-scale clinical settings. Since our method is capable of processing direct-stained smears, only basic staining supplies are required for slide preparation. Rapid staining kits such as the QBC Diagnostics F.A.S.T. kit are viable in field settings and could thus be used with CellScope in remote areas that lack laboratory infrastructure.

Previous Work. The two main methods of screening sputum samples are fluorescence microscopy (FM) and brightfield microscopy, in which the sputum smears are stained with auramine-O and Ziehl-Neelsen respectively (see Figure 2). CellScope is capable of both types of microscopy, but we focus on FM here because studies indicate it is more sensitive and significantly faster [3,13]. Several groups have explored automated TB detection for conventional FM microscopes. Veropoulos *et al.* [18] applied Canny edge detection, filtered objects based on size, and used boundary tracing to identify candidate objects. Fourier descriptors, intensity features, and compactness were then combined with various probabilistic classification methods, and a multilayer neural network achieved the best performance. Forero *et al.* [9] took a generative approach, representing the TB-bacilli class with a Gaussian mixture model (GMM) and using Bayesian classification techniques. Hu moment features were chosen for their invariance to rotation, scaling, and translation. Other groups have proposed algorithms for brightfield microscopy [7,12], but these algorithms often rely on the distinct color characteristics of Ziehl-Neelsen staining.
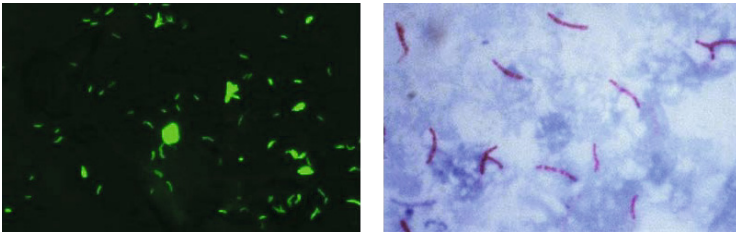
Fig. 2. *Left:* Sample CellScope fluorescence image. *Right:* Sample brightfield image [14].

Additional TB diagnostic procedures include culture and polymerase chain reaction (PCR)-based methods. Culture results are ideally used to verify smear screenings and are the current gold-standard for diagnosis. However, culture assays are more expensive and technically challenging to perform than smear microscopy and require prolonged incubation: about 2-6 weeks to allow accurate evaluation of bacteria. PCR-based methods such as Cepheid's GeneXpert assess the presence of TB bacterial DNA and are rapid, more sensitive than smear microscopy, and capable of testing resistance to a common anti-TB antibiotic [1]. However, PCR-based methods continue to lag in sensitivity compared to culture and rely on costly equipment that is poorly suited for low-resource, peripheral healthcare settings [8]. Sputum smear microscopy continues to be by far the most widely used method of TB diagnosis, suggesting that enhancements to microscopy-based screening methods could provide significant benefit to large numbers of TB-burdened communities across the globe.

2   Methods and Materials

2.1   Algorithm

We propose a TB detection algorithm for FM with three stages: (1) candidate TB-object identification, (2) feature representation, and (3) discriminative classification. A block diagram of the algorithm is shown in Figure 3.

Candidate TB-Object Identification. In the first stage, our goal is to identify any bright object that is potentially a TB-bacillus. We perform a white top-hat transform and template matching with a Gaussian kernel. The white top-hat transform reduces noise from fluctuations in the background staining, and the template matching picks out areas that resemble bright spots. The result is a binarized image, from which we extract the connected components as candidates. We consider a region of interest or patch from the input image centered around each candidate. The patch-size (24x24 pixels) is chosen based on the known size of the TB-bacilli (typically 2-4μm in length and 0.5μm in width) and CellScope's sample-referenced pixel spacing of 0.25μm/pixel.

Feature Representation. We characterize each candidate TB-object using Hu moments [11]; geometric and photometric features; and histograms of oriented
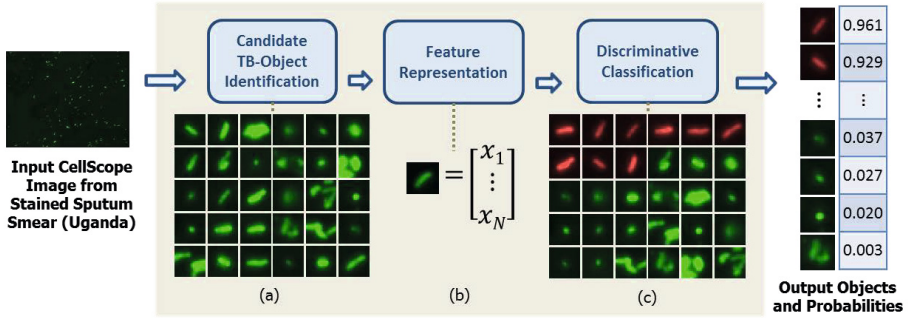
Fig. 3. Overview of algorithm. (a) Array of candidate TB-objects. (b) Each candidate characterized by 102-dimensional feature vector. (c) Candidates sorted by decreasing probability of being a TB-bacillus (row-wise, top to bottom). Sample subset of candidate TB-objects with corresponding probabilities shown at the output. Object-level probabilities subsequently used to determine slide-level diagnosis.

gradients (HOG) [4]. Hu moments, photometric features, and HOG are calculated from the grayscale patch, whereas geometric properties are determined from a binarized version of the image patch. Binarization is achieved using Otsu's method [15], which minimizes the variance within each of the two resulting pixel classes. Eight Hu moment features provide a succinct object-level description that is invariant to rotation, translation, and scaling (similar to [9]). In addition, we calculate fourteen geometric and photometric descriptors: area, convex area, eccentricity, equivalent diameter, extent, filled area, major/minor axis length, max/min/mean intensity, perimeter, solidity, and Euler number. Finally, we extract HOG features from each 24x24 patch using two scales and 8 orientations, giving eighty HOG feature values. We thus obtain a 102-dimensional feature vector representing the appearance of each candidate TB-object.

Candidate TB-Object Classification. We consider three object-level classifiers in our experiments (in order of increasing discriminative power and computational cost): logistic regression, linear support vector machines (SVMs) and intersection kernel (IK) SVMs [5,6,17]. Intuitively, SVMs find the hyperplane that maximizes the margin between the TB-positive and TB-negative classes in the feature space. IKSVMs achieve nonlinear decision boundaries via the intersection kernel, defined as $K(\mathbf{u}, \mathbf{v}) = \sum_i min(\mathbf{u}_i, \mathbf{v}_i)$. We normalize the input feature vectors using maximum-minimum standardization and apply logistic regression to the SVM outputs to obtain probabilities [16], which indicate the likelihood of each object being a TB-bacillus.

Performance Metrics. We present our experimental results using two sets of performance metrics: Recall/Precision and Sensitivity/Specificity, which are widely used in the computer vision and medical communities respectively. Recall refers to the fraction of true positive objects correctly classified as positives,

while Precision refers to the fraction of objects classified as positive that are true positives. Sensitivity is the same as Recall, and Specificity is Recall for the negative class. Recall/Precision are more appropriate for gauging object-level performance in this study because our negative class is much larger than our positive class. At the slide level, however, our data has balanced class sizes and thus both Recall/Precision and Sensitivity/Specificity are suitable. In this study, we optimize over Average Precision (AP) at the slide level, which places equal weight on Recall and Precision. Often in practice it is more useful to have *either* very high Precision *or* very high Recall (rule-in or rule-out value, respectively) rather than moderately high values for both. In these cases, one may instead optimize over the maximum F_β-measure, defined as $F_\beta = (1 + \beta^2)\frac{\text{Precision}\cdot\text{Recall}}{(\beta^2\cdot\text{Precision})+\text{Recall}}$, where $\beta < 1$ gives more weight to Precision than Recall ($\beta = 1$ gives equal weight).

2.2   Dataset and Ground Truth

Our dataset consists of sputum smear slides collected at clinics in Uganda. Fluorescence images of these smears were taken using CellScope, which has a 0.4NA objective and an 8-bit monochrome CMOS camera. CellScope gives a Rayleigh resolution of $0.76\mu m$ and is capable of effective magnifications of 2000-3000x. The CellScope images are 1944x2592 pixels and cover a $640x490\mu m$ field of view at the smear-referenced plane. We use 594 CellScope images (296 TB-positive, 298 TB-negative), which correspond to 290 patients (143 TB-positive, 147 TB-negative). We have slide-level human reader and culture classification results for all 290 slides. In addition, a human annotator labeled TB-objects in a subset of the positive images (92 of 296 images), resulting in 1597 positive TB-objects. The human readers in this study received guidance from experts, and their performance has been shown to be statistically comparable to that of trained microscopists. Our dataset and human annotations will be publicly available.

3   Experimental Results and Discussion

Object-Level Evaluation. For the object-level classification task, we use the subset of TB-positive images for which we have human annotations and all TB-negative images. Applying our object identification procedure, we retain 98.8% of the positive TB-objects in the dataset after the first step. All objects identified in TB-negative images are considered negative objects. This results in 1597 positive and 34948 negative objects, which correspond to 390 images (92 positive and 298 negative).

We generate five random training-test splits with our object-level data: one for model parameter selection and four to assess robustness of results. We train various object-level classifiers, using slide-level performance as the optimizing criterion for parameter selection. We then perform systematic ablation studies as summarized in Figure 4. We find that the best performance is achieved when using the whole feature set with an IKSVM: Average Precision of $89.2\% \pm 2.1\%$
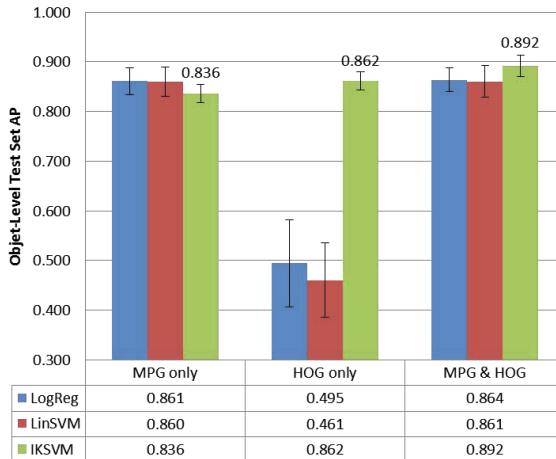
Fig. 4. Object-level test set AP across different classifiers (logistic regression, linear SVM, and IKSVM) and feature subsets. Two categories of features: Hu moments/photometric/geometric (MPG) and histograms of oriented gradients (HOG).

over the four remaining test sets. When relying solely on HOG features, logistic regression and linear SVM methods perform poorly. This is expected because the HOG features are not rotation invariant. [18] also evaluated their algorithm performance at the object-level, but their data and implementation code are not publicly available for direct comparison.

Slide-Level Evaluation. We also consider algorithm performance at the slide level, which is more relevant for practical diagnosis. Because slide-level culture results are available, evaluating our algorithm at the slide level frees us from human-labeled ground truth. To determine slide-level decisions from object-level scores, we refer to how experts manually classify slides. For each slide, we gather the output SVM scores of all the objects and average the top K scores, where $K = 3$ is chosen via validation experiments. We classify the *slide* as positive if the averaged score falls above a given threshold. By varying this threshold, we obtain a Recall-Precision curve (see plot in Figure 5). As shown in Figure 5, we consider the three object-level classifiers (logistic regression, linear SVMs, and IKSVMs) in terms of their slide-level performance. We adopt the IKSVM because it achieves slightly better slide-level performance than the other two methods. On the four remaining test sets, the IKSVM achieves slide-level Average Precision of $92.3\% \pm 0.9\%$ and Average Specificity of $88.0\% \pm 1.3\%$.

Slide-Level Comparison with Baseline and Human Readers. We compare our algorithm's slide-level performance to that of human readers and Forero's GMM-based approach [9]. We train Forero's algorithm using our data, where color filtering is reduced to intensity filtering because CellScope images are monochromatic. The GMM method achieves Average Precision of $79.7\% \pm 3.3\%$ and

Method	AP(%)	Max F_1-meas(%)
Humans	-	85.9±1.3
Our SVM	92.3±0.9	84.9±2.4
Baseline	79.7±3.3	78.8±1.8

Classifier	AP(%)	AS(%)
LogReg	91.4±0.5	87.1±1.2
LinSVM	91.1±1.2	86.4±1.3
IKSVM	92.3±0.9	88.0±1.3
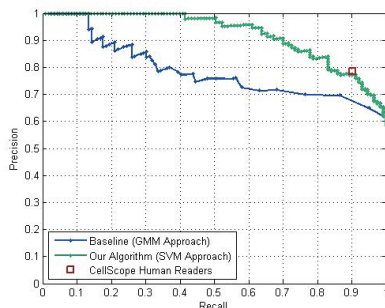
Fig. 5. Slide-Level Performance. *Left top:* Comparison of our IKSVM-based algorithm's performance to that of humans and the baseline method (GMM approach). Average Precision (AP) and maximum F_1-measure across four test sets. *Right:* Slide-level Recall-Precision curves across different methods for one test set. *Left bottom:* Our algorithm's slide-level performance for different object-level classifiers. Average Precision (AP) and Average Specificity (AS), where we average over four test sets.

maximum F_1-measure of 78.8%±1.8% (see Figure 5). Human readers also inspected the same CellScope images and classified each slide, resulting in an F_1-measure of 85.9%±1.3% across the four test sets. The plot in Figure 5 shows Recall/Precision curves across different methods for a sample training-test split. For that split, we see that our algorithm's slide-level performance is comparable to that of human readers and achieves a higher fraction of true positives than the GMM approach for most Recall values.

4   Summary and Conclusions

We propose an accurate and robust automated TB detection algorithm for low-cost, portable digital microscopes such as the CellScope. Applying modern computer vision techniques to images from mobile microscopes could save lives in low-resource communities burdened by TB and suffering poor access to high-quality TB diagnostics. The sputum smears used in our study were collected in Uganda and provide a realistic dataset for algorithm training and evaluation. Our algorithm first identifies potential TB-objects and characterizes each candidate object using Hu moments, geometric and photometric features, and histograms of oriented gradients. We then classify each of the candidate objects using an IKSVM, achieving Average Precision of 89.2% ± 2.1% for object classification. At the slide level, our algorithm performs as well as human readers, showing promise for making a tremendous impact on global TB healthcare. We will release our dataset, annotations, and code, which we hope will provide helpful insights for future approaches to quantitative TB diagnosis.

References

1. Boehme, C.C., Nabeta, P., Hillemann, D., Nicol, M.P., Shenai, S., Krapp, F., Allen, J., Tahirli, R., Blakemore, R., Rustomjee, R., Milovic, A., Jones, M., O'Brien, S.M., Persing, D.H., Ruesch-Gerdes, S., Gotuzzo, E., Rodrigues, C., Alland, D., Perkins, M.D.: Rapid Molecular Detection of Tuberculosis and Rifampin Resistance. New England J. of Med. 363(11), 1005–1015 (2010)
2. Breslauer, D.N., Maamari, R.N., Switz, N.A., Lam, W.A., Fletcher, D.A.: Mobile Phone Based Clinical Microscopy for Global Health Applications. PLoS One 4(7), e6320 (2009)
3. Cattamanchi, A., Davis, J.L., Worodria, W., den Boon, S., Yoo, S., Matovu, J., Kiidha, J., Nankya, F., Kyeyune, R., Byanyima, P., Andama, A., Joloba, M., Osmond, D.H., Hopewell, P.C., Huang, L.: Sensitivity and Specificity of Fluorescence Microscopy for Diagnosing Pulmonary Tuberculosis in a High HIV Prevalence Setting. Int. J. Tuberc. Lung Dis. 13(9), 1130–1136 (2010)
4. Dalal, N., Triggs, B.: Histograms of Oriented Gradients for Human Detection. In: CVPR, pp. 886–893 (2005)
5. Chang, C.C., Lin, C.J.: LIBSVM: A Library for Support Vector Machines. ACM Trans. on Intell. Sys. and Tech. 2(3), 27:1–27:27 (2011)
6. Cortes, C., Vapnik, V.: Support-vector networks. Mach. Learn. 20(3), 273–297 (1995)
7. Costa, M.G., Costa Filho, C.F., Sena, J.F., Salem, J., de Lima, M.O.: Automatic Identification of Mycobacterium Tuberculosis with Conventional Light Microscopy. In: 30th Ann. Int. IEEE EMBS Conf., pp. 382–385 (2008)
8. Evans, C.A.: GeneXpert-A Game-Changer for Tuberculosis Control? PLoS Med. 8, e1001064 (2011)
9. Forero, M.G., Cristóbal, G., Desco, M.: Automatic Identification of Mycobacterium Tuberculosis by Gaussian Mixture Models. J. of Microscopy 223(2), 120–132 (2006)
10. Global Tuberculosis Control: WHO Report 2011 (2011), http://www.who.int/tb/publications/global_report/
11. Hu, M.K.: Visual Pattern Recognition by Moment Invariants. IRE Trans. on Info. Theory 8(2), 179–187 (1962)
12. Khutlang, R., Krishnan, S., Dendere, R., Whitelaw, A., Veropoulos, K., Learmonth, G., Douglas, T.S.: Classification of Mycobacterium Tuberculosis in Images of ZN-Stained Sputum Smears. IEEE Trans. on Info. Tech. in Biomed. 14(4), 949–957 (2010)
13. Kivihya-Ndugga, L.E.A., van Cleeff, M.R.A., Githui, W.A., Nganga, L.W., Kibuga, D.K., Odhiambo, J.A., Klatser, P.R.: A Comprehensive Comparison of Ziehl-Neelsen and Fluorescence Microscopy for the Diagnosis of Tuberculosis in a Resource-Poor Urban Setting. Int. J. Tuberc. Lung Dis. 7(12), 1163–1171 (2003)
14. Kubica, G.P.: Mycobacterium Tuberculosis Bacteria Using Acid-Fast Ziehl-Neelsen Stain; magnified 1000x. Public Health Image Library, Centers for Disease Control and Prevention, Atlanta (1979)
15. Otsu, N.: A Threshold Selection Method from Gray-Level Histograms. IEEE Trans. on Systems, Man and Cybernetics 9(1), 62–66 (1979)
16. Platt, J.C.: Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods. In: Advances in Large Margin Classifiers, pp. 61–74 (1999)
17. Fan, R.E., Chang, K.W., Hsieh, C.J., Wang, X.R., Lin, C.J.: LIBLINEAR: A Library for Large Linear Classification. JMLR 9, 1871–1874 (2008)
18. Veropoulos, K.: Machine learning approaches to medical decision making. U. of Bristol (2001)

Accurate Fully Automatic Femur Segmentation in Pelvic Radiographs Using Regression Voting

C. Lindner[1], S. Thiagarajah[2], J.M. Wilkinson[2], arcOGEN Consortium, G.A. Wallis[3], and Timothy F. Cootes[1]

[1] Imaging Sciences, University of Manchester, UK
[2] Department of Human Metabolism, University of Sheffield, UK
[3] Wellcome Trust Centre for Cell Matrix Research, University of Manchester, UK

Abstract. Extraction of bone contours from radiographs plays an important role in disease diagnosis, pre-operative planning, and treatment analysis. We present a fully automatic method to accurately segment the proximal femur in anteroposterior pelvic radiographs. A number of candidate positions are produced by a global search with a detector. Each is then refined using a statistical shape model together with local detectors for each model point. Both global and local models use Random Forest regression to vote for the optimal positions, leading to robust and accurate results. The performance of the system is evaluated using a set of 519 images. We show that the fully automated system is able to achieve a mean point-to-curve error of less than $1mm$ for 98% of all 519 images. To the best of our knowledge, this is the most accurate automatic method for segmenting the proximal femur in radiographs yet reported.

Keywords: automatic femur segmentation, femur detection, Random Forests, Hough Transform, Constrained Local Models, radiographs.

1 Introduction

In clinical practice, plain film radiographs are widely used to assist in disease diagnosis, pre-operative planning and treatment analysis. Extraction of the contours of the proximal femur from anteroposterior (AP) pelvic radiographs plays an important role in diseases such as osteoarthritis (e.g. diagnostics and joint-replacement planning) or osteoporosis (e.g. fracture detection and bone density measurements). In addition, accurately segmenting the contours of the proximal femur in radiographs allows monitoring of disease progression.

Manual segmentation of the femur is time-consuming and hard to do consistently. Our aim is to automate the segmentation procedure. Fully automatic proximal femur segmentation is challenging for several reasons: *(i)* The quality of radiographs may vary a lot in terms of contrast, resolution and the region of the pelvis shown. *(ii)* AP pelvic radiographs only give a 2D projection, and hence are susceptible to rotational issues; the same 3D shape may yield a different 2D projection depending on the view point. *(iii)* Plain film radiographs do not provide homogeneous values for the same structure due to overlapping body parts.

(iv) Deformities of the proximal femur may cause the loss of distinguishable radiographic key features.

Automatically extracting the contours of the proximal femur comprises two key steps: Firstly, the femur is detected in the image and secondly, the contours are segmented. Behiels et al. [1] have shown the suitability of statistical shape models for proximal femur segmentation. Recent work on automatically segmenting the femur in radiographs using statistical shape models includes [11,12]. Object detection in the latter as well as in the atlas-based approach of Ding et al. [8] is based on edge detection. We use Random Forest regression in a sliding window approach to automatically segment the proximal femur.

Random Forests (RF) [2] describe an ensemble of decision trees trained independently on a randomised selection of features. They have been shown to be effective in a range of classification and regression problems [6,10]. Recent work on Hough Forests [9] has shown that objects can be effectively located by training RF regressors to predict the position of a point relative to the sampled region, then running the regressors over a region and accumulating votes for the likely position. To detect the femur, our global search uses a RF regressor that votes for the centre of a reference frame, resulting in a response image of accumulated votes. The approximated position is then used to initialise a local search to segment the femur, combining local detectors with a statistical shape model. Following [3], we apply RF regression in the Constrained Local Model (CLM) framework to vote for the optimal position of each model point. Here, feature detectors are run independently to generate response images for each point and then a shape model is used to find the best combination of points [7].

Using RF regression voting for both object detection and CLM-based contour extraction yields a robust and fully automatic segmentation system. We use the latter to segment the femur in pelvic radiographs, and demonstrate that results are very accurate. The local search and the fully automatic search outperform alternative matching techniques such as Active Shape Models [5], CLMs using normalised correlation and RF classification-based search. We believe this to be the most accurate fully automatic femur segmentation system yet published.

2   Methods

The fully automated segmentation system comprises a global search detecting the object and a local search segmenting the contours. Both global and local search use RF regression voting to predict object and point positions.

2.1   Voting with Random Forest Regression

We use RF regression in a similar manner to the Hough Forests approach [9]. However, we do not require voting to be dependent on a class label, allowing all image structures to vote. In the voting-regression approach, we evaluate a set of points in a grid over a region of interest. At each point $\mathbf{z}$, a set of features $\mathbf{f}(\mathbf{z})$ is sampled. A regressor, $R(\mathbf{f}(\mathbf{z}))$, is trained to predict the most likely position(s) of

the target point relative to **z**. During training, given the samples at a particular node, we seek to select a feature and threshold to best split the data. Let f_i be the value of one feature associated with sample i. The best threshold, t, for this feature at this node is the one which minimises $G_T(t) = G(\{\mathbf{d}_i : f_i < t\}) + G(\{\mathbf{d}_i : f_i >= t\})$ where $G(S)$ is a function evaluating the set of vectors S, and $\mathbf{d}_i$ the predicted displacement of sample i. We aim at minimising the entropy in the branches when splitting the nodes using $G(\{\mathbf{d}_i\}) = N log|\Sigma|$, where N is the number of displacements in $\{\mathbf{d}_i\}$ and Σ the respective covariance matrix. Criminisi et al. [6] showed that a related measure of information gain was effective for regression.

Hough Forests use RFs whose leaves store multiple training samples. Thus each sample produces multiple votes, allowing for arbitrary distributions to be encoded. Each leaf of our decision trees only stores the mean offset and the standard deviation of the displacements of all training samples that arrived at that leaf. During search, these predictions are used to vote for the best position in an accumulator array. Predictions are made using a single vote per tree yielding a Hough-like response image. To blur out impulse responses we slightly smooth the response image with a Gaussian.

Below, we use Haar features [13] as they have been found to be effective for a range of applications and can be calculated efficiently from integral images.

2.2   Object Detection

Training. A reference frame, or bounding box, is set to capture the object of interest. For each training image, a number of random displacements (scale, angle and position) of the bounding box are sampled. To train the detector, for every sample we extract features $\mathbf{f}_i$ at a set of random positions within the sampled patch and store displacement $\mathbf{d}_i$ from the original centre of the reference frame. We then train a RF on the pairs $\{\mathbf{f}_i, \mathbf{d}_i\}$. To train a single tree, we take a bootstrap sample of the training set, and construct the tree by recursively splitting the data at each node as described in Section 2.1. The extracted features are a random subset of all possible Haar features and at each node, we choose the feature and associated threshold which minimise G_T to split the data.

Search. To detect the object in an image, we scan the image at a set of coarse angles and scales in a sliding window approach. The search is speeded up by evaluating only positions on a sparse grid rather than at every pixel. For every angle-scale combination, we scan the bounding box across the image. We obtain the relevant feature values from each box and get the RF to make predictions on the reference frame centre. Predictions are made using a single Gaussian weighted vote per tree, where the weights relate to the spread of the displacements of the training samples that arrived at the particular leaf. The resulting response image is then searched for local maxima. Once a response image has been obtained for every angle-scale combination, all maxima are ranked according to their total votes. Every maxima is associated with an angle, a scale and a prediction of the reference frame centre. This results in candidate positions for the object.

2.3   Segmentation Using Constrained Local Models

CLMs combine global shape constraints with local models of pattern of intensities. Based on a number of landmark points outlining the contour of the object in a set of images, we train a statistical shape model by applying PCA to the aligned shapes [5]. This yields a linear model of shape variation which represents the position of each landmark point using $\mathbf{x}_i = T_\theta(\bar{\mathbf{x}}_i + \mathbf{P}_i\mathbf{b} + \mathbf{r})$ where $\bar{\mathbf{x}}_i$ gives the mean in the reference frame, $\mathbf{P}_i$ is a set of modes of variation, $\mathbf{b}$ are the shape model parameters, $\mathbf{r}$ allows small deviations from the model, and T_θ applies a global transformation (e. g. similarity) with parameters θ. Similar to Active Appearance Models [4], CLMs combine this shape model with a texture model but only sample a local patch around each landmark rather than the whole object.

To match the CLM to a new image, we seek the shape and pose parameters, $\mathbf{p} = \{\mathbf{b}, \theta\}$, which optimise the fit of the model to the image. Given an initial estimate of every landmark's position, an area around each landmark point is searched. At every position i, a quality-of-fit value, describing the similarity between the template texture for this landmark learned from the model and the texture at that position, is obtained and stored in a response image $\mathbf{R}_i$. We then find the shape and pose parameters which optimise $\Sigma_{i=1}^n \mathbf{R}_i(T_\theta(\bar{\mathbf{x}}_i + \mathbf{P}_i\mathbf{b} + \mathbf{r}))$.

In [3] it is shown how RF regression voting produces useful response images for the CLM framework. Here we summarise the key steps.

Training. CLMs in their original form use normalised correlation as quality-of-fit measurement for each response image. In the RF regression approach, we train a regressor to predict the position of a landmark point based on a random set of Haar features. The quality-of-fit values here relate to the votes of the RF.

For every landmark i, we sample local patches at a number of random displacements $\mathbf{d}_i$ from the true position. For every sample we extract features $\mathbf{f}_i$ and train a RF on the pairs $\{\mathbf{f}_i, \mathbf{d}_i\}$. As with the global search, we train every tree taking a bootstrap sample and constructing it recursively by splitting the data at each node as described in Section 2.1.

Search. To match the RF regression-based CLM to a new image, for every landmark i, we sparsely sample local patches in the area around an initial estimate of the landmark's position. We extract the relevant features for each sample and get the RF to make predictions on the true position of the landmark. Predictions are made using a single vote per tree. This yields a response image $\mathbf{R}_i$ for every landmark i. We then aim to combine voting peaks in the response images with the global constraints learned by the shape model.

2.4   Automated System

The fully automated system performs a global search at multiple scales and orientations to produce a number of candidate poses which are ranked by total votes. The local search is then applied at each of the best l search candidates, and the final results are ranked by the total CLM fit (sums of votes).

3   Experiments and Evaluation

The aim is to fully automatically segment the femur by putting a dense annotation of 65 landmarks along its contours as demonstrated in Figure 1; (a) gives the manual annotation and (b)-(c) the result of the fully automated system. We use a *front-view femur* model that excludes both trochanters and approximates the superior lateral edge (points 43 to 47) from an anterior perspective. All points were defined using anatomical features mixed with an evenly spaced subset.

(a)   (b)   (c)
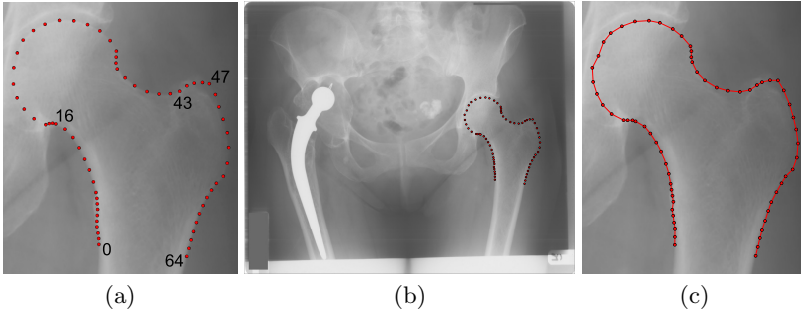
Fig. 1. Segmentation of the proximal femur: (a) 65 landmarks outlining the 'front-view' femur (ground truth); (b)-(c) automatically segmented femur in AP pelvic radiograph

Our data set comprises AP pelvic radiographs of 519 females suffering from unilateral hip osteoarthritis. All images were provided by the arcOGEN Consortium and were collected under relevant ethical approvals. The images have been collected from different radiographic centres resulting in varying resolution levels (555-4723 pixels wide) and large intensity differences. In addition, the displayed pelvic region and the pose of the femur in the images vary a lot. For each image, a manual annotation of 65 landmarks as in Figure 1(a) is available. In the following we performed two-fold cross-validation experiments, averaging results from training on each half of the data and testing on the other half.

3.1   Global Search: Automatic Femur Detection

We set up a detector that samples the whole proximal femur and three regions of interest (shaft, femoral head, greater trochanter). For each of the latter, we train a RF of 10 trees using samples at 20 random pose and scale displacements.

During search, the object detector scans the image at a range of coarse orientations and scales, and provides the 40 best fits. Each match determines candidate positions for points 16 and 43 (see Figure 1), defining a reference length. All candidates are clustered using a cluster radius of 10% of the reference length. We evaluate the mean point-to-point error as a percentage of the reference length, and give results for the *best* (minimal mean error) cluster only. When averaging over both reference points, the detector yields an error of less than 11.4%

for 95% of all 519 images. Our data set contains 15 calibrated images suggesting an average reference length of $57mm$. Using the latter, the error of the global search relates to less than $6.5mm$ for 95% of all images.

3.2   Local Search: Accurate Femur Segmentation

We train a RF regression-based CLM using a reference frame that is 200 pixels wide and a patch size of 15x15 pixels within the reference frame. For each training image and every landmark, we sample 20 patches using random displacements of up to 20 pixels in x and y in the reference image, as well as random displacements in scale ($\pm 5\%$) and rotation ($\pm 6°$). We train a RF of 10 trees for every landmark.

To compare the performance of the RF regression-based CLM with alternative techniques, we train a correlation-based CLM and a RF classification-based CLM using the same settings, as well as an ASM. All models are trained to explain 95% of the shape variation given by the training set, and start searching from the mean shape at the correct pose. Figure 2(a) shows the mean point-to-curve error as a percentage of the shaft width. We define the latter as the distance between landmarks 0 and 64 (see Figure 1). We use this as a reference length as it tends to be relatively constant across individuals; our calibrated subset suggests an average length of $37mm$. Results show that the RF regression-based CLM performs best with a mean point-to-curve error of within 2.0% for 95% of all images, which relates to a local search accuracy of within $0.7mm$.
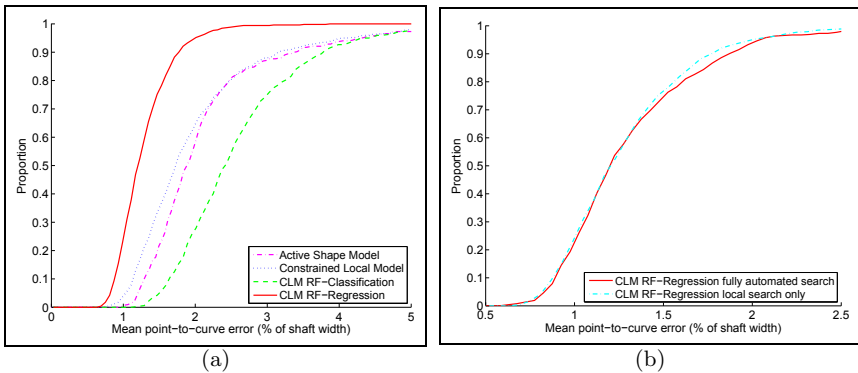
Fig. 2. Quantitative evaluation: (a) local search results starting from mean shape at true pose; (b) fully automated search showing results for the best clustered candidate

3.3   Full Search: Accurate Automatic Femur Segmentation

For the fully automated system, we use the clustered candidates obtained via the global search to initialise the local search. Every candidate predicts the positions of points 16 and 43. This initialises the scale and pose of the RF regression-based CLM. We test all candidates for every image, and run 20 search iterations from

the initialised mean model. We choose the candidate that gives the best final quality-of-fit value to give the fully automatic segmentation result.

Figure 2(b) gives the mean point-to-curve error of the fully automated system as a percentage of the shaft width. This shows that the global search works sufficiently well for the fully automated system to be very accurate with errors of less than 2.1% for 95% of all 519 images, relating to $0.8mm$. The overlapping plots indicate that the fully automated system yields almost equally high accuracy as a local search starting from the mean shape at the correct pose.

Figure 3 shows various segmentation results of the fully automated system, ranked according to mean point-to-curve percentiles: (a) gives the median result (50% of the images have a mean error of less than $0.5mm$); (b) is based on the second highest global search error yielding a mean segmentation error of $0.7mm$; (c)-(d) show the two highest mean segmentation errors where (c) achieved an accuracy of $1.6mm$ and (d) is the only case out of 519 images where the global search failed to initialise the local search sufficiently well.

(a)   (b)   (c)   (d)
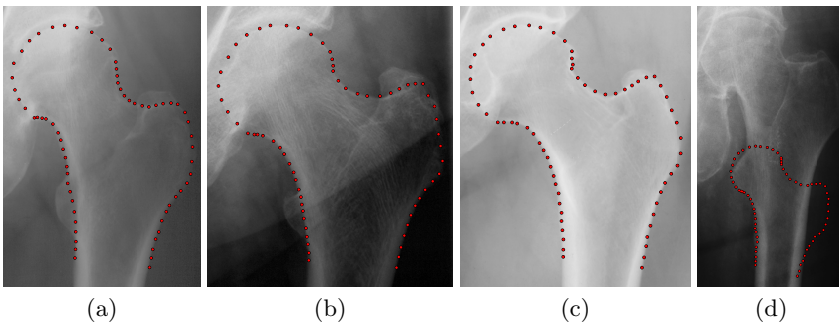
Fig. 3. Examples of segmentation results of the fully automated system (sorted by the mean point-to-curve percentiles): (a) median; (b) 92.1%, based on second highest global search error; (c) 99.8%, second highest overall error; (d) maximal overall error, only example where global search failed to sufficiently initialise the local search. (Due to space we only show the proximal femur; all searches were run on full pelvic images.)

A direct comparison to other reported results seems difficult as most findings are either given qualitatively, or are not easy to interpret in more general terms. The best reported results appear to be the ones by Pilgram et al. [11] with a point-to-curve error of within $1.6mm$ for 80% of the 117 test cases (estimated on the basis of likely shaft width relative to image width).

4   Discussion and Conclusions

We have presented a system to segment the proximal femur in AP pelvic radiographs which is fully automatic, does not make any assumptions about the femur pose, and is very accurate. We have shown that the system achieves excellent performance when tested on a set of 519 images of mixed quality. The femur detector, achieving an accuracy of a mean point-to-point error of less than $8.4mm$

for 99% of all images, works generally sufficiently well to initialise the local model used for segmentation. In our experiments, the fully automatic segmentation system achieved an overall mean point-to-curve error of less than $1mm$ for 98% of all images. We believe that this is the most accurate fully automatic system for segmenting the proximal femur in AP pelvic radiographs so far reported.

All experiments were run on a 3.3 GHz Intel Core Duo PC using 2GB RAM. The global search took on average 15s per image, and the local search 10s per image and cluster; we searched on average 10 clusters. Note that running times vary depending on image size and search settings. The fully automated system is sufficiently general to be applied to other medical segmentation problems.

References

1. Behiels, G., Maes, F., Vandermeulen, D., Suetens, P.: Evaluation of image features and search strategies for segmentation of bone structures in radiographs using Active Shape Models. Medical Image Analysis 6(1), 47–62 (2002)
2. Breiman, L.: Random forests. Machine Learning 45, 5–32 (2001)
3. Cootes, T., Ionita, M., Lindner, C., Sauer, P.: Robust and accurate shape model fitting using random forest regression. Tech. Rep. 2012-01, Uni. Manchester (2012)
4. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active Appearance Models. In: Burkhardt, H., Neumann, B. (eds.) ECCV 1998. LNCS, vol. 1407, pp. 484–498. Springer, Heidelberg (1998)
5. 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)
6. Criminisi, A., Shotton, J., Robertson, D., Konukoglu, E.: Regression Forests for Efficient Anatomy Detection and Localization in CT Studies. In: Menze, B., Langs, G., Tu, Z., Criminisi, A. (eds.) MICCAI 2010 Workshop MCV. LNCS, vol. 6533, pp. 106–117. Springer, Heidelberg (2011)
7. Cristinacce, D., Cootes, T.: Automatic feature localisation with Constrained Local Models. Journal of Pattern Recognition 41(10), 3054–3067 (2008)
8. Ding, F., Leow, W.-K., Howe, T.S.: Automatic Segmentation of Femur Bones in Anterior-Posterior Pelvis X-Ray Images. In: Kropatsch, W.G., Kampel, M., Hanbury, A. (eds.) CAIP 2007. LNCS, vol. 4673, pp. 205–212. Springer, Heidelberg (2007)
9. Gall, J., Lempitsky, V.: Class-specific Hough forests for object detection. In: CVPR, pp. 1022–1029. IEEE Press (2009)
10. Girshick, R., Shotton, J., Kohli, P., Criminisi, A., Fitzgibbon, A.: Efficient regression of general-activity human poses from depth images. In: ICCV, pp. 415–422. IEEE Press (2011)
11. Pilgram, R., et al.: Knowledge-based femur detection in conventional radiographs of the pelvis. Computers in Biology and Medicine 38, 535–544 (2008)
12. Smith, R., Najarian, K., Ward, K.: A hierarchical method based on active shape models and directed Hough transform for segmentation of noisy biomedical images. BMC Medical Informatics and Decision Making 9(suppl. 1), 2–12 (2009)
13. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: CVPR, pp. 511–518. IEEE Press (2001)

Automatic Location of Vertebrae on DXA Images Using Random Forest Regression⋆

M.G. Roberts, Timothy F. Cootes, and J.E. Adams

Imaging Science Research Group, University of Manchester, U.K.
`martin.roberts@manchester.ac.uk`

Abstract. We provide a fully automatic method of segmenting verte-
brae in DXA images. This is of clinical relevance to the diagnosis of
osteoporosis by vertebral fracture, and to grading fractures in clinical
trials. In order to locate the vertebrae we train detectors for the upper
and lower vertebral endplates. Each detector uses random forest regres-
sor voting applied to Haar-like input features. The regressors are applied
at a grid of points across the image, and each tree votes for an endplate
centre position. Modes in the smoothed vote image are endplate candi-
dates, some of which are the neighbouring vertebrae of the one sought.
The ambiguity is resolved by applying geometric constraints to the con-
nections between vertebrae, although there can be some ambiguity about
where the sequence starts (e.g. is the lowest vertebra L4 or L5, Fig 2a).
The endplate centres are used to initialise a final phase of Active Ap-
pearance Model search for a detailed solution. The method is applied
to a dataset of 320 DXA images. Accuracy is comparable to manually
initialised AAM segmentation in 91% of images, but multiple grade 3
fractures can cause some edge confusion in severely osteoporotic cases.

1 Introduction

The accurate identification of vertebral fractures is clinically important in the
diagnosis of osteoporosis. Typically diagnosis uses a semi-quantitative approach
using subjective judgement by a radiologist. Quantitative morphometric meth-
ods are not specific and require tedious manual annotation of six points on each
vertebra. See [1] for a discusson of diagnosis methods. More sophisticated clas-
sification methods based on statistical models have been reported in [1,2], but
these require an accurate segmentation method. Active appearance models [3]
(AAM) have been used to segment dual energy X-ray absorptiometry (DXA)
images in [4], but the method required a manual initialisation on the centre
of each vertebra. In applications such as clinical drug trials, it is desirable to
eliminate the manual initialisation.

This paper describes a three-phase approach. First we locate putative verte-
bral endplates using a set of random forest regressors (one per endplate), together
with Hough-style tree voting. Modes in the vote image of each vertebral end-
plate are candidate endplate positions. Secondly, the ambiguity is resolved by

---

⋆ We are thankful to Arthritis Research UK for funding.

applying a graphical model of the connections between vertebrae, thus applying geometric constraints. The solution of the graph problem is used to initialise a third phase of AAM search. The novelty of the work lies in providing a fully automatic segmentation method for vertebrae, although there still remains an inevitable ambiguity in the starting level (e.g. is the bottom vertebra detected in Fig 2a L4 or L5), which even experienced clinicians can find difficult to resolve.

1.1 Data

The dataset used consists of 320 DXA Vertebral Fracture Assessment (VFA) images scanned on various Hologic (Bedford MA) scanners, obtained from: a) 44 patients from a previous study [5]; b) 80 female subjects in a epidemiological study of a UK cohort born in 1946; c) 196 females attending a local clinic for DXA BMD measurement, and for whom the referring physician had also requested VFA (as approved by the local ethics committee). The lumbar vertebrae from L4-L1 and the thoracic vertebrae from T12-T7 (Fig 2c) were annotated with detailed point positions (42 points per vertebra) for training AAMs.

2 Methods

2.1 Regression Forests

Background on Regression Trees. Regression trees [6] are an efficient way of predicting continuous output vectors given a complex set of input features. At each branch in the tree the data is split into two subsets based on a threshold on a selected feature using some criterion that seeks to increase the homogeneity of the output vectors in the child branches. A regression forest consists of multiple trees with some degree of randomisation, for example each tree is trained on a bootstrapped subset; and at each branch a random subset of the input features are considered. After training, a new data point can be predicted by dropping the input vector down the trees. Each tree produces a prediction, which is typically the mean of the training outputs at the terminal node. The predictions of each tree in the forest are then aggregated, typically by taking the mean over all trees; or in [7] the distributions of each leaf node are used; or Hough style voting can be used [8]. Unlike Hough forests [8] we do not train object classification; each of our trees cast one vote; and we include votes cast by patches which may be displaced completely outside the object.

Pre-processing and Training. The images are smoothed with a Gaussian filter with $\sigma =$1mm for L4-L1 and then reduced according to mean vertebral size in the thoracic spine. The image is locally normalised using mean and variance derived from a sliding exponential filter of standard deviation 25mm. Patches are then sub-sampled with a step size of 1.5σ based on the appropriate endplate centre and aligned to the vertebral axis. For the lumbar vertebrae (L4-L1) the patch extends from the centre of the endplate 18 mm inside the vertebra and

12mm outside, whilst in width it extends to the mean semi-width across the training set, plus a left distance of 18mm and 12mm to the right. The left (posterior) bias is to include further context information from the spine. These distances are downscaled for the smaller thoracic vertebrae. Similar patches are also sampled by randomly displacing the centre up to 30mm in x, 20mm in y (with an x-axis aligned to the true vertebral axis), and randomly rotating by up to $20°$. The (axis-relative) displacements in x, y and orientation are stored after normalising.

Input Features and Training. The input features used are all possible Haar-like features [9]. In [7] the features used were the difference in mean intentsity values between two randomly displaced boxes. We also experimented with these, but found that the Haar features performed slightly more accurately and reliably. We believe that broadly similar results can be obtained from random box-comparison features, but the Haar scheme includes more complex comparisons for highlighting ridges and corners.

In order to randomise between trees, and reduce the large Haar set, at each tree branch we pick a random subset of features (mean size 1000). We generated 100 perturbations for each patch in each image and trained the random forest regressors using leave-20-out cross-validation. The splitting criteria used was the total variance (weighted by sub-sample size) summed over the output variables. The selected splitting feature and threshold are those that minimise the weighted total variance, subject to an imposed minimum node sample size of 7. A branch was terminated if the node variance reached a minimum variance of 1% of the total initial variance, or at a depth of 18. We used a forest size of 40 trees. The output is the 3-dimensional vector giving the displacements in x, y (axis-relative) and rotation.

Voting Scheme. Endplate centre candidates are located by sampling the Haar features across a grid of points separated by 4mm in x and 2mm in y, and at a set of 5 orientations $\{\alpha_r^{(t)}\}$, given by the mean vertebral axis and displacements between $-20°$ and $+20°$ at $10°$ intervals. The grid is sampled over the 3 SD range of positions for that vertebra in the training set together with a 10% bounding region. We initially maintain separate voting accumulators for each of the 5 starting orientations. Each grid position casts votes from each tree in the forest as follows. For tree i starting at grid position $\mathbf{x}_j$ for endplate r, the angular displacement prediction $\hat{\theta}_{ijr}$ is used to update the vertebral axis $\hat{\alpha}_{ijr} = \alpha_r^{(t)} + \hat{\theta}_{ijr}$; then after applying the counter-rotation $\mathbf{R}^T(\hat{\alpha}_{ijr})$ from the tree's axis-relative frame, we obtain the predicted location $\hat{\mathbf{x}}_{ijr} = \mathbf{x}_j + \mathbf{R}^T[\Delta x_{ijr}, \Delta y_{ijr}]^T$. This update scheme also using orientation prediction is more complex than simply predicting $\Delta x, \Delta y$ in a world frame, but since the features are defined relative to a vertebral axis, it should be more accurate; it also allows us to introduce orientation weighting (see below) in the voting scheme.

The predicted location $\hat{\mathbf{x}}_{ijr}$ is used to accumulate a vote array in the neighbourhood of the rounded position in an array of 1x1mm^2 bins, using Gaussian

kernel smoothing with $\sigma = 0.5$ (half a bin). The vote portion accumulated in bin position $\mathbf{b}_{lm}$ is $w_{lm}/\sum_{(l,m)\in N} w_{lm}$ with N the neighbourhood such that $|\mathbf{b}_{lm} - \hat{\mathbf{x}}_{ijr}| \leq 2$, and $w_{lm} = \exp(-|\mathbf{b}_{lm} - \hat{\mathbf{x}}_{ijr}|^2/2\sigma^2)$.

The vote of each tree is further weighted by angular displacement $w_a(\hat{\theta}_{ijr})$, as we expect better accuracy for starting orientations closely aligned to the real orientation; where

$$w_a(\hat{\theta}_{ijr}) = 1 \quad \left|\hat{\theta}_{ijr}\right| <= \theta_0; \qquad w_a(\hat{\theta}_{ijr}) = \exp\left(-\frac{|\hat{\theta}_{ijr}|-\theta_0}{\sigma_\theta}\right) \quad \left|\hat{\theta}_{ijr}\right| > \theta_0$$

We use $\theta_0 = 5°$ and $\sigma_\theta = 10°$. The total vote for each accumulator at each point V_{lmr} is obtained by summing across all starting grid points and trees so $V_{lmr} = \sum_i \sum_j w_a(\hat{\theta}_{ijr})w_{lm}(\hat{\mathbf{x}}_{ijr})$. Each vote accumulator is then treated as a 2D image and further smoothed with SD 1.5mm. All local modes in this smoothed vote image $\check{\mathbf{V}}$ are located, and the top 20 modes from each orientation are passed on for clustering.

The clustering algorithm forms the minimum spanning tree of all modes pooled from all orientations, and then deletes all arcs of length exceeding 2mm. The remaining connected sub-graphs form clusters and the feature location is taken as the highest scoring mode in the cluster. The vote score $\check{V}_{kr}$ of mode k is post-processed onto a $(0,1)$ scale to form a quasi-probability $\check{p}_{kr}$, using a sigmoidal transform parameterised using the inter-quartile range of the successfully located modes in the training set. This transform can be viewed as a biased version of the logistic function commonly used in classification [1], or as an approximation to the CDF of the successful mode vote. The top 10 modes for each endplate location give a candidate set of scores $\{\check{p}_r\}$ and positions $\{\check{\mathbf{x}}_r\}$.

2.2   Inter-Feature Geometric Constraints

Typically the correct endplate position is somewhere in the list of modes, but similar responses are encountered at neighbouring vertebrae (Fig 2a), and the lower endplate detectors can locate neighbouring upper endplates or vice versa (Fig 2b). To resolve the ambiguity, we use a geometric model containing multiple nodes (one per endplate), together with a model of the pairwise geometrical
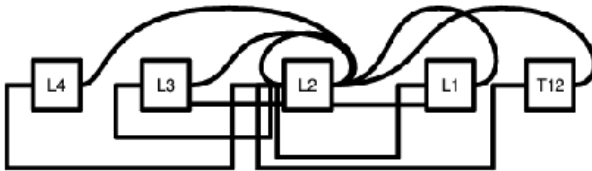
Fig. 1. The arcs connecting L2 endplates to neighbours. Similar connections exist for other vertebrae.

---

[1] Note that the graph solution (see below) is invariant to constant bias.

relationships between them as in [10]. We can then use graph algorithms to locate the optimal solution for the combination of feature response and geometry. We connect the lower endplate to the upper endplate, which is also connected to the lower endplate immediately above it. This chain can be solved by dynamic programming. However a more complex graph is needed to allow for missing detection cases, and also additional constraints can help resolve ambiguities. So arcs also join each upper/lower endplate to the corresponding endplates of vertebrae above and below including two vertebrae distant (Fig 1). The arc costs are given by the Mahalanobis distance D_{rq} of each pair in the edge set E after aligning into the target vertebral axis frame. The final selection of optimal solution $\tilde{\mathbf{k}}$ from the set of candidate modes $\{\check{\mathbf{x}}_{rk}\}$ is given by finding the minimum sum of node and arc costs over the R endplates:

$$\tilde{\mathbf{k}} = \mathrm{argmin}_{\mathbf{k}} \left\{ \sum_{r=1}^{R} -\log(\check{p}_r(k_r)) + \lambda \sum_{(r,q) \in E} D_{rq}(k_r, k_q) \right\} \qquad (1)$$

The graph is not a tree and cannot be solved by dynamic programming, but we use loopy belief propagation (LBP) instead [11] [2]. The parameter λ controls the relative weighting on the spatial constraints; we used $\lambda = 0.1$, based on some provisional experiments with related random forest classifiers (i.e. is a box centred on the endplate or not). We found the detection of T8/T7 to be somewhat unreliable (may be obscured by the scapulae, Fig 2d), and so solve the graph problem from L4-T9, and leave T8/T7 to the next phase of AAM fitting.

2.3   Active Appearance Models

We train AAMs [3] for overlapping triplets of vertebrae similar to [4] covering L4-T7. We initialise a global shape model using the endplate centres using a robust M-estimator to allow for some detection failures. Then a set of individual AAMs are initialised for each triplet covering L4 to T9. Initially we concentrate on the more reliable L4-T11 section, and at each iteration perform a tentative fit to all remaining triplets, and then pick the best one (lowest residual sum of squares in AAM texture model) to impose. This affects the re-initialisation of the neighbours. After fitting L4-T11, the remaining triplets are fitted by moving up the spine, and after each triplet fit a global shape model is updated to predict the positions of the remaining vertebrae used in their AAM initialisation.

2.4   Experiments

We tested the algorithms using leave-20-out cross-validation on the 320 images, and calculated point-to-line errors against the gold standard manual annotation. There is a fundamental ambiguity in determining the vertebral levels. The lowest

---

[2] The max-product variant, equivalent to max-sum with log probabilities.

visible vertebra may be L4 or L5 (even L3) (Fig 2a); false positives can be detected on the sacrum; or there can even be an L6. The optimum graph solution can correspond to the correct solution shifted up or down by one vertebra (occasionally even two). We have not yet addressed how to resolve this ambiguity. In order to produce meaningful overall accuracy statistics we examine the top solution, and other solutions after removing the lowest nodes in the former, up to 5 possible solutions. If any of these identify the vertebral centres L4-T12 to within 4mm in Y and 6mm in X we consider it a success, and proceed with the highest ranked such solution to the AAM stage.
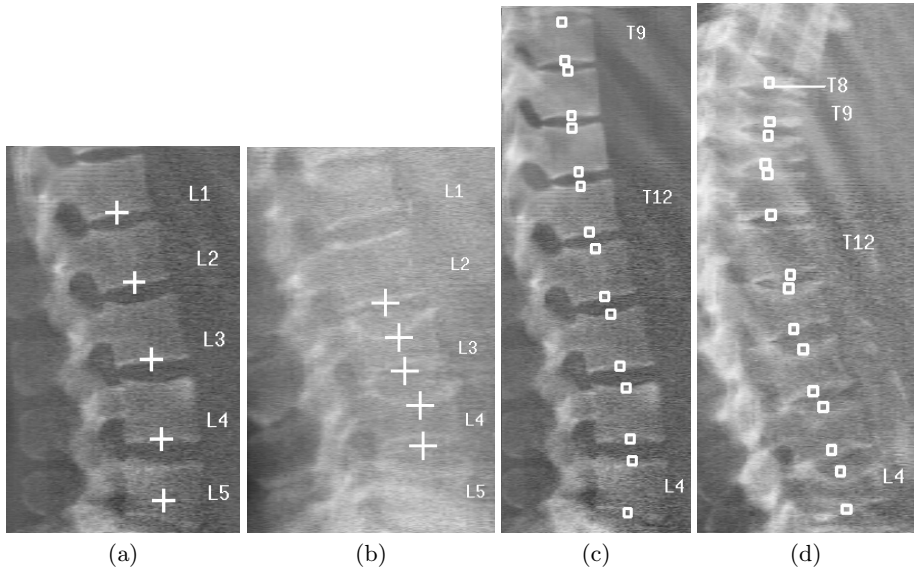
Fig. 2. Detected positions for bottom of L2 and overall LBP graph solutions used to initialise AAM. a) Top 5 L2 (bottom) regressor mode positions on bottoms of L5-L1; b) Similar regressor modes on bottoms of L2-L4 plus confounders at the tops of L3, L4; note L1,L3,L4 fractures; c) Good image with accurate LBP solution for all vertebrae; d) Severely osteoporotic case - most fractures are located by the LBP solution but the top of T9 is mis-located on the bottom of the extremely fractured T8 (indicated).

3   Results

Figure 2 (a-b) shows two examples of the top 5 located positions for the L2 lower endplate. These typically include responses from neighbours, but there are responses on the upper endplates which can produce some confusing shifts. Figure 2 (c-d) also shows examples of the solution for L4-T9 used to initialise the AAM. A successful initialisation was obtained in 308 images out of 320 (96.2%). The failures were mostly severely osteoporotic cases with many fractures (hence unusual geometries), or severe disc disease fusing vertebrae. In the 308 successful

Table 1. Search error statistics (point-to-line) for AAM by vertebra fracture status. Bracketed figures are after excluding the 15 images (5%) where 3 or more vertebrae suffered edge confusion with neighbours.

Vertebra Status	%ge of Sample	Automatic Initialisation Search Error Statistic			Manual Initialisation Search Error Statistic		
		Mean (mm)	Median (mm)	%ge errors >2mm	Mean (mm)	Median (mm)	%ge errors >2mm
Normal	84.9%	0.66(0.55)	0.41(0.40)	3.6(2.5)%	0.55(0.55)	0.40(0.40)	2.5(2.3)%
Grade 1	5.9%	0.92(0.75)	0.52(0.50)	7.7(5.4)%	0.70(0.70)	0.49(0.49)	4.8(4.8)%)
Grade 2	5.1%	1.12(0.88)	0.61(0.58)	11.4(9.3)%	0.92(0.88)	0.61(0.59)	10.2(9.3)%
Grade 3	4.1%	1.94(1.18)	0.80(0.67)	23.9(15.0)%	1.19(1.07)	0.72(0.68)	16.5(14.0)%

images, 60% have the best solution at the correct vertebral level, with 27% and 13% shifted up or down by one vertebra respectively, and a single case is shifted by two.

Table 1 gives point-to-line accuracy results for these 308 successful cases, together with corresponding figures for a manually initialised AAM search. In these 308 cases the overall mean segmentation error was 0.74mm, increasing with fracture grade. After running the AAM there were 15 images (5%) with edge confusions on 3 or more vertebrae in succession - typically caused by successive severe fractures (such as Fig 2d). Removing these images from the statistics substantially reduces the mean error, which is skewed by the larger error tail on these partial failures. The errors on the images without these substantial edge confusions are given in brackets in table 1, which also shows errors from manually initialised AAM fits for comparison. If the 15 edge confusion images are also considered failures, then the overall success rate is 91.2%, and in these cases the accuracy is comparable to that obtained using a manual initialisation (overall mean 0.59mm vs 0.58mm manual).

4   Discussion and Conclusions

Although there is some failure of the automatic initialisation process, the algorithm sucessfully locates a plausible set of vertebrae in over 91% of cases, with most failures on extremely osteoporotic cases. The fundamental ambiguity of vertebral levels is still an unsolved problem, but in some triage applications (e.g. detect any patient with possible fracture) this may not matter; or the user can be presented with the best solution plus two shifted versions to choose from. Good accuracy is obtained for normal (unfractured) vertebrae, but there are larger errors for severely fractured vertebrae, due to edge confusions between a vertebra and its neighbours. The overall mean error of 0.74mm compares well to other methods (e.g. [12], mean error 1.4mm, or [13], mean errors 1.22 or 1.34mm). Our initialisation failure of 3.8% on L4-T9 appears comparable to the 2% on L4-L1 for the baseline data in [13], when scaled by the number of vertebrae. The final failure rate of 8.8% appears higher, but [13] deals only with the lumbar, whereas

our additional failures are due to multiple severe fractures in the thoracic spine. The thoracic vertebrae are harder to segment because the vertebrae are closer together (especially when affected by disc disease), resulting in more edge confusions between neighbouring vertebrae; and there is more overlaying structure from the ribs and scapulae. If the algorithm were being used in triage to pick up patients with any fractures, then even the additional 15 image "failures" would still result in successful patient referrals, as all are cases like Fig 2d with some fractures being successfully located before edge confusion occurs.

References

1. Roberts, M.G., Cootes, T.F., Pacheco, E.M., Adams, J.E.: Quantitative vertebral fracture detection on DXA images using shape and appearance models. Academic Radiology 14, 1166–1178 (2007)
2. de Bruijne, M., Lund, M., Tanko, L., Pettersen, P., Nielsen, M.: Quantitative vertebral morphometry using neighbour-conditional shape models. Med. Image Anal. 11, 503–512 (2007)
3. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active appearance models. IEEE Transactions on Pattern Analysis and Machine Intelligence 23, 681–685 (2001)
4. Roberts, M.G., Cootes, T.F., Adams, J.E.: Vertebral morphometry: semi-automatic determination of detailed shape from DXA images using active appearance models. Investigative Radiology 41(12), 849–859 (2006)
5. McCloskey, E.V., et al.: Effects of clodronate on vertebral fracture risk in osteoporosis: a 1-year interim analysis. Bone 28(3), 310–315 (2001)
6. Breiman, L.: Random forests. Machine Learning 45, 5–32 (2001)
7. Criminisi, A., Shotton, J., Robertson, D., Konukoglu, E.: Regression Forests for Efficient Anatomy Detection and Localization in CT Studies. In: Menze, B., Langs, G., Tu, Z., Criminisi, A. (eds.) MICCAI 2010 Workshop MCV. LNCS, vol. 6533, pp. 106–117. Springer, Heidelberg (2011)
8. Gall, J., Lempitsky, V.: Class-specific hough forests for object detection. In: Proc of CVPR 2009, pp. 1022–1029. IEEE Computer Society (2009)
9. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Proc of CVPR 2001, pp. 511–518. IEEE Computer Society (2001)
10. Donner, R., Micusik, B., Langs, G., Bischof, H.: Sparse MRF appearance models for fast anatomical structure localisation. In: Rajpoot, N., Bhalerao, A. (eds.) Proc. of BMVC 2007, pp. 1080–1089. BMVA (2007)
11. Weiss, Y., Freeman, W.: On the optimality of solutions of the max-product belief propagation algorithm in arbitrary graphs. IEEE Trans. Inf. Theory 47, 736–744 (2001)
12. de Bruijne, M., Nielsen, M.: Image segmentation by shape particle filtering. In: Proc. of ICPR 2004, pp. 722–725. IEEE Computer Society (2004)
13. Petersen, K., Ganz, M., Mysling, P., Nielsen, M., Lillemark, L., Crimi, A., Brandt, S.: A bayesian framework for automated cardiovascular risk scoring on standard lumbar radiographs. IEEE Trans. Med. Imag. 31(3), 663–676 (2012)

Decision Forests for Tissue-Specific Segmentation of High-Grade Gliomas in Multi-channel MR

Darko Zikic[1], Ben Glocker[1], Ender Konukoglu[1], Antonio Criminisi[1],
C. Demiralp[2], J. Shotton[1], O.M. Thomas[3,4], T. Das[3], R. Jena[3],
and S.J. Price[3,5]

[1] Microsoft Research Cambridge, UK
[2] Brown University, Providence, RI, USA
[3] Cambridge University Hospitals, Cambridge, UK
[4] Department of Radiology, Cambridge University, UK
[5] Department of Clinical Neurosciences, Cambridge University, UK

Abstract. We present a method for automatic segmentation of high-grade gliomas and their subregions from multi-channel MR images. Besides segmenting the gross tumor, we also differentiate between active cells, necrotic core, and edema. Our discriminative approach is based on decision forests using context-aware spatial features, and integrates a generative model of tissue appearance, by using the probabilities obtained by tissue-specific Gaussian mixture models as additional input for the forest. Our method classifies the individual tissue types simultaneously, which has the potential to simplify the classification task. The approach is computationally efficient and of low model complexity. The validation is performed on a labeled database of 40 multi-channel MR images, including DTI. We assess the effects of using DTI, and varying the amount of training data. Our segmentation results are highly accurate, and compare favorably to the state of the art.

1 Introduction

In this paper, we present our work on tissue-specific segmentation of high-grade gliomas in multi-channel MR images, with focus on grade IV glioblastoma tumors. Such high-grade gliomas (HGG) grow rapidly, infiltrate the brain in an irregular way, and often create extensive vasculature networks. HGGs contain a necrotic core (NC), surrounded by a varyingly thick layer of active cells (AC). Together, necrotic core and active cells form the gross tumor (GT). Usually, the tumor itself is surrounded by a varying amount of edema (E). In consequence, HGGs have extremely heterogeneous shape, appearance and location (cf. Figs. 1,2), which makes their automatic analysis challenging.

Our goal is to segment high-grade gliomas as well as the individual tissue components automatically and reliably. This would 1) speed-up accurate delineation of the tissue components, which is crucial for radiotherapy and surgery planning and is currently performed manually in a labor intensive fashion, and 2) allow direct volume measurements. Volume measurements are critical for the evaluation

of treatment [17], however seldom performed since manual tumor segmentation is often impractical in a routine clinical setting. Instead more straightforward but less accurate measures are used, such as a pair of perpendicular tumor diameters [17]. Distinguishing between volumes of individual tissue types, especially active cells and necrotic core is an important step for assessment of treatment response. For example, an effective drug might not change the gross tumor volume, while transforming active into necrotic cells. To detect this change, the volumes of both these tissues must be monitored.

This paper proposes an efficient method for automatic segmentation of glioblastoma in multi-channel MR images. While most of the previous research focuses on segmentation of gross tumor, or tumor and edema, we perform a tissue-specific segmentation of three relevant tissues types: active cells (AC), necrotic core (NC), and edema (E). Our method is based on decision forests (DF) [3], a discriminative model which we combine with a generative model of tissue appearance. We achieve this by using the probability estimates based on Gaussian mixture models (GMM) as additional input for the forest. An important advantage of DFs is that they are inherently multi-label classifiers, which allows us to classify the different tissues simultaneously. Such simultaneous separation of classes has the potential to simplify the modeling of the distributions of the individual classes. Through the use of context-sensitive features in the forest, our approach yields a natural smoothness of the segmentation results without explicit regularization. Our method has a low model complexity and reduces the necessity for a large number of pre- and post-processing steps.

The accuracy of our method is evaluated quantitatively on a database of 40 high-grade glioma patients – to our knowledge, the largest annotated database of this kind so far – and compares favorably to the results in the literature.

1.1   Related Work

In recent years several approaches for segmentation of brain tumors have been proposed. The settings differ from one another in many respects, such as the type of tumor being handled (e.g. low-grade [6], high-grade [4,13], pediatric tumors [16]), the type of anomalous tissues being detected (e.g. GT [7,10,16]; GT and E [4,6,12,15]; AC, NC, E [2,15]), input data, and the evaluation procedure.

A popular group of methods is based on the registration of patient images to a probabilistic brain atlas [6,9,10,12]. The main idea is that – given an accurate atlas and registration – the tumor can be detected as deviation of patient data from the atlas. Since the presence of the tumor makes the registration challenging, some approaches use manual interaction [9], while others integrate tumor growth models [6].

A large group of discriminative methods applies learning techniques to the problem [2,4,7,13,15,16]. Our method belongs to this group. Mostly, a learning method is combined with a regularization step, e.g. by modeling the boundary [8,11], or by applying a variant of a random field spatial prior (MRF/CRF) [4,7,16]. Works which classify multiple labels [2,15] often use SVMs, which are inherently *binary* classifiers. In order to classify different tissues, they are applied
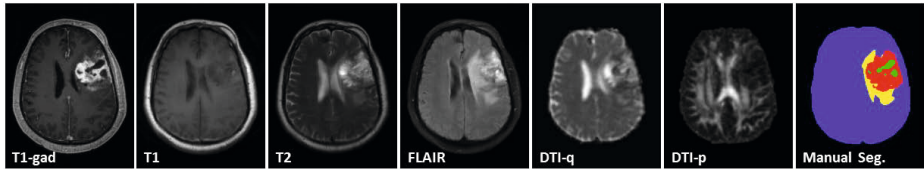
Fig. 1. Example of one of 40 patients in our high-grade glioma database, with tissues labeled as active cells (red), necrotic core (green), and edema (yellow)

hierarchically [2], or in the one-versus-all manner [15]. For these approaches, several classes have to be grouped together, a step which can make the distribution inside the aggregate group more complex than the distribution of each individual class. For example, the intensity distribution of a tumor consisting of AC and NC tissues, which have very different representations in the multichannel data, is likely to be more complex than the distributions of the single classes. In our technique, we circumvent this potential problem by classifying all tissues simultaneously, which allows us to only handle distributions of individual classes.

Finally, discriminative methods are sometimes seen as requiring heavy data processing and mandatory spatial regularization [10]. In our discriminative approach, despite using only minimal amounts of pre-processing, we achieve high accuracy results without post-hoc regularization.

2   The Labeled High-Grade Glioma Database

We acquired a set of multi-channel 3D MR data for 40 patients suffering from high-grade gliomas, with 38 cases of grade IV tumors (glioblastomas) and 2 grade III tumors (anaplastic oligodendrogliomas). The data is acquired prior to treatment. For each patient we have the following 6 channels: T1 post gadolinium (T1-gad), T1, T2 turbo spin echo (T2-tse), and FLAIR, and 2 channels from diffusion tensor imaging (DTI-p and DTI-q). Fig. 1 gives an example for one patient. All acquisitions were performed on a 3T Siemens TrioTim. We will refer to the multi-channel data as I_{MR}. For all 40 patients, a manual segmentation of the three classes of AC, NC, and E is obtained in 3D (see Figs. 1,2).

We try to keep the amount of data pre-processing at a minimum. We perform skull stripping of MR channels [14], and for each patient we perform an affine intra-patient registration of all channels to the T1-gad image. No inter-patient registration is required. We also avoid a full bias-field correction, and only align the mean intensities of the images within each channel by a global multiplicative factor. All these steps are fully automatic.

3   Method: Decision Forests with Initial Probabilities

In our approach we use decision forests (DF)[3,5] as a discriminative multiclass classification method, and combine them with a generative model of tissue

appearance. This is achieved by using initial tissue probability estimates based on trained GMMs as additional input channels for the forest, along with the MR data I_{MR}. We classify four classes AC, NC, and E, and background (B), and gross tumor remains defined as GT=AC∪NC.

As the first step of our approach, we estimate the initial class probabilities for a given patient as posterior probabilities based on the likelihoods obtained by training a set of GMMs on the training data. For each class c, we train a single GMM, which captures the likelihood $p(I_{\mathrm{MR}}|c)$ of the multi-dimensional intensity for this class. For a given patient data set I_{MR}, the GMM-based posterior probability p_c^{GMM} for the class c is estimated for each point $x \in \mathbb{R}^3$ by $p^{\mathrm{GMM}}(c|x) = p(I_{\mathrm{MR}}(x)|c)\,p_c\,/\,\sum_{c_j} p(I_{\mathrm{MR}}(x)|c_j)\,p_{c_j}$, with p_c denoting the prior probability for the class c, based on its relative frequency. We can now use the probabilities $p_c^{\mathrm{GMM}}(x)=p^{\mathrm{GMM}}(c|x)$ directly as input for the decision forests, in addition to the multi-channel MR data. So now, our data for one patient is a set of n_{c} channels $I=$(T1-gad, T1, T2, FLAIR, DTI-q, DTI-p, $p_{\mathrm{AC}}^{\mathrm{GMM}}, p_{\mathrm{NC}}^{\mathrm{GMM}}, p_{\mathrm{E}}^{\mathrm{GMM}}, p_{\mathrm{B}}^{\mathrm{GMM}}$).

For simplicity, we denote single channels by I_j, and the data for a patient k by $I^{(k)}$. Please note that we can use the GMM-based probabilities for maximum a posteriori classification by $\hat{c} = \arg\max_c p^{\mathrm{GMM}}(c|x)$. We will use this for a base line comparison in Sec. 4.

3.1   Decision Forests

We employ decision forests (DF) to determine a class $c \in C$ for a given spatial input point $x \in \Omega$, based on the representation of x by a feature vector $f(x, I)$. DFs are ensembles of (binary) decision trees, indexed by $t \in [1, n]$. As a supervised method, DFs operate in two stages: training and testing.

During training, each tree t learns a weak class predictor $p_t(c|f(x, I))$ for a n_{f}-dimensional feature representation $f(x, I) \in \mathbb{R}^{n_{\mathrm{f}}}$ of a spatial point x from the data set I. The input training data set is $\{(f(x, I^{(k)}), c^{(k)}(x)) : x \in \Omega^{(k)}\}$, that is, the feature representations of all spatial points $x \in \Omega^{(k)}$, in all training patient data sets k, and the corresponding manual labels $c^{(k)}(x)$. We refer to all spatial points in all training data sets by $X=\bigcup_k \Omega^{(k)}$. We will use $x \in X$ to identify single training examples in most part, thus writing e.g. $p_t(c|x)$ for $p_t(c|f(x, I))$.

In a decision tree, each node i contains a set of training examples X_i, and a class predictor $p_t^i(c|x)$, which is the probability corresponding to the fraction of points with class c in X_i. Starting with X at the root, the training is performed by successively splitting the training examples at every node based on their feature representation, and assigning the partitions X_{L} and X_{R} to the left and right child node. At each node, a number of splits along randomly chosen dimensions of the feature space is considered, and the one maximizing the Information Gain is applied. Tree growing is stopped at a certain tree depth d.

At testing, a point x to be classified is pushed through each tree t, by applying the learned split functions. Upon arriving at a leaf node l, the leaf probability is used as the tree probability, i.e. $p_t(c|x)=p_t^l(c|x)$. The overall probability is computed as the average of tree probabilities, i.e. $p(c|x)=\frac{1}{n}\sum_{t=1}^n p_t(c|x)$. The actual class estimate $\hat{c}$ is chosen as the most probable class, i.e. $\hat{c} = \arg\max_c p(c|x)$.
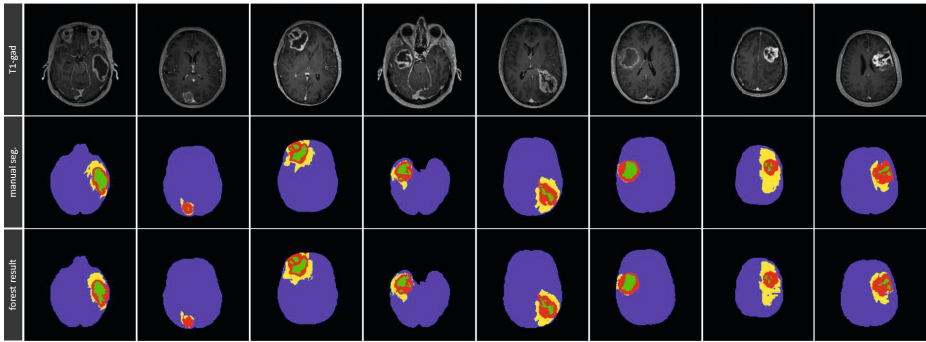
Fig. 2. Examples of results on 8 patients. Obtained by a forest with GMM, MR, and DTI input, with training on 30 patients. The high accuracy of our results is quantitatively confirmed in Figs. 3,4 (AC=red, NC=green, E=yellow).

3.2 Context-Aware Feature Types

We employ three spatial and context-aware features types, which are intensity-based and parametrized. Two of these feature types are generic [5], while the third one is designed with the intuition of detecting structure changes. Every instantiated feature with its unique parameters corresponds to one dimension of the feature space used by decision trees.

We use the following notation: Again, $x \in \Omega$ is a spatial point, to be assigned a class, and I_j is an input channel. $N_j^s(x)$ denotes an x-centered and axis aligned 3D cuboid in I_j with edge lengths $s=(s_x, s_y, s_z)$, and $v \in \mathbb{R}^3$ is an offset vector.

Feature Type 1: Intensity difference between x in a channel I_{j_1} and an offset point $x + v$ in a channel I_{j_2} (note that $I_{j_1}=I_{j_2}$ is allowed)

$$f^1(x, I)_{j_1, j_2, v} = I_{j_1}(x) - I_{j_2}(x + v) \; . \tag{1}$$

Feature Type 2: Difference between intensity means of a cuboid around x in I_{j_1}, and around an offset point $x + v$ in I_{j_2}

$$f^2(x, I)_{j_1, j_2, s_1, s_2, v} = \mu(N_{j_1}^{s_1}(x)) - \mu(N_{j_2}^{s_2}(x + v)) \; . \tag{2}$$

Feature Type 3: Intensity range along a 3D line between x and $x+v$ in one channel. This type is designed with the intuition that structure changes can yield a large intensity change, e.g. NC being dark and AC bright in T1-gad.

$$f^3(x, I)_{j, v} = \max_{\lambda}(I_j(x + \lambda v)) - \min_{\lambda}(I_j(x + \lambda v)) \quad \text{with} \quad \lambda \in [0, 1] \; . \tag{3}$$

4 Evaluation

We perform an extensive series of cross-validation experiments to evaluate our method. For this, the 40 patients are randomly split into non-overlapping training and testing data sets. To investigate the influence of the size of the training
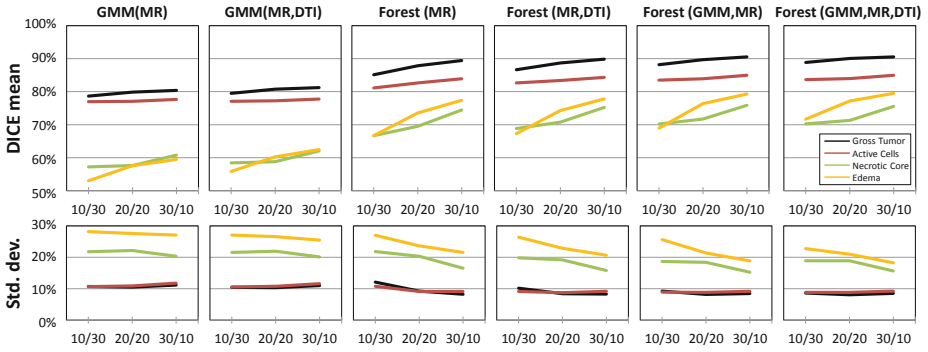
Fig. 3. Average mean and standard deviations of DICE scores, for experiments on 10 random folds, with the training/testing data set sizes of 10/30, 20/20, and 30/10. From left to right, the approaches yield higher mean scores, with lower std. devs. Our approach (rightmost) shows increased robustness to amount of training data, resulting in more horizontal lines, indicating better generalization.

set and generalization properties of our method, we perform experiments with following training/testing sizes: 10/30, 20/20, 30/10. For each of the three ratios, we perform 10 tests, by randomly generating 10 different training/testing splits.

To demonstrate the influence of the single components of the method, we also perform tests on Forests without GMMs, and compare to the results of GMM only. Finally, we investigate the influence of using DTI, by performing all experiments also with MR input only. Overall, this results in 30 random training sets, and 600 tests for each of the 6 approaches. The evaluation is performed with all images sampled to isotropic spatial resolution of 2mm, and forests with $n=40$ trees of depth $d=20$. With these settings, the training of one tree takes between 10-25 min, and testing 2-3 min, depending on the size of training set and the number of channels. The algorithm and feature design were done on a single independent 20/20-fold, which was not used for evaluation.

Fig. 2 shows a visual example of the results, while the quantitative evaluation and more details are given in Figs. 3,4. We observe an improvement of the segmentation accuracy by the proposed method (Forest(GMM,MR,DTI)) compared to the other tested configurations. The amount of training data influences NC and E more than AC and GT. The effect of using DTI seems to be most visible for smaller training data sets.

Comparison to Quantitative Results of Other Approaches is difficult for a number of reasons, most prominently the different input data. To provide some indicative context, we cite results of a recent work from [2]. There, the mean and standard deviation for a leave-one-out cross-validation on 10 glioma patients, based on multichannel MR are as follows: GT: 77±9, AC: 64±13, NC: 45±23, E: 60±16. Our results compare favorably. For our 30/10-tests we get: GT: 90±9, AC: 85±9, NC: 75±16, E: 80±18, and for the more challenging 10/30-tests (less training data), we get GT: 89±9, AC: 84±9, NC: 70±19, E: 72±23.
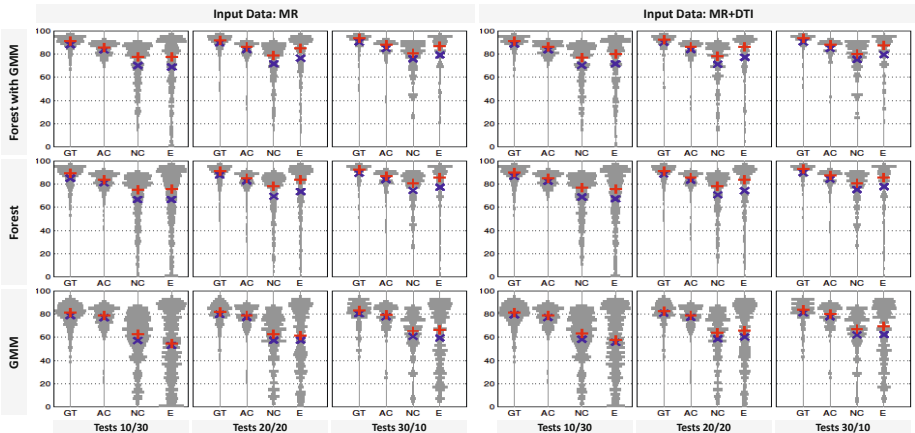
Fig. 4. Evaluation of 10 random fold experiments with varying amount of training data. Distribution of DICE scores is indicated along y-axis, by plotting the histogram of the scores, grouped per test and tissue, with medians (+) and means (×). Our forest method clearly reduces the number of lower score outliers compared to GMM. The outliers occur mostly for NC and E, and visual inspection confirms that the misclassification of NC and E is the most significant error of our method. Increasing the amount of training data reduces the number of outliers. DTI has most effect for less training data, and the GMM estimates.

Sensitivity to Variation of Parameters is tested by varying $n \in [15, 40]$ and $d \in [12, 20]$, for the ten 30/10-tests. We observe robustness to the selection of these values, especially n. Details are given in the supplementary material [1].

5 Summary and Conclusion

We propose a method for automatic and tissue-specific segmentation of high-grade gliomas. Our discriminative approach is based on decision forests using context-aware features, integrates a generative model of tissue appearance, and classifies different tissues simultaneously. Our method requires comparably little pre-processing, and no explicit regularization, thus resulting in a low model complexity. The approach is computationally efficient, reasonably robust to parameter settings, and achieves highly accurate segmentation results. The automatic results are suitable for volume measurements, and can be used as high-quality initial estimates for interactive treatment planning.

References

1. Sup. material, http://research.microsoft.com/apps/pubs/default.aspx?id=164382

2. Bauer, S., Nolte, L.-P., Reyes, M.: Fully Automatic Segmentation of Brain Tumor Images Using Support Vector Machine Classification in Combination with Hierarchical Conditional Random Field Regularization. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part III. LNCS, vol. 6893, pp. 354–361. Springer, Heidelberg (2011)
3. Breiman, L.: Random forests. Machine Learning 45(1), 5–32 (2001)
4. 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 Trans. Medical Imaging 27(5) (2008)
5. Criminisi, A., Shotton, J., Konukoglu, E.: Decision forests: A unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning. FnT Computer Graphics and Vision (2012)
6. Gooya, A., Pohl, K.M., Bilello, M., Biros, G., Davatzikos, C.: Joint Segmentation and Deformable Registration of Brain Scans Guided by a Tumor Growth Model. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part II. LNCS, vol. 6892, pp. 532–540. Springer, Heidelberg (2011)
7. Görlitz, L., Menze, B.H., Weber, M.-A., Kelm, B.M., Hamprecht, F.A.: Semi-supervised Tumor Detection in Magnetic Resonance Spectroscopic Images Using Discriminative Random Fields. In: Hamprecht, F.A., Schnörr, C., Jähne, B. (eds.) DAGM 2007. LNCS, vol. 4713, pp. 224–233. Springer, Heidelberg (2007)
8. Ho, S., Bullitt, E., Gerig, G.: Level-set evolution with region competition: automatic 3-D segmentation of brain tumors. In: ICPR (2002)
9. Kaus, M.R., Warfield, S.K., Nabavi, A., Black, P.M., Jolesz, F.A., Kikinis, R.: Automated segmentation of brain tumors. Radiology 218 (2001)
10. Menze, B.H., van Leemput, K., Lashkari, D., Weber, M.-A., Ayache, N., Golland, P.: A Generative Model for Brain Tumor Segmentation in Multi-Modal Images. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part II. LNCS, vol. 6362, pp. 151–159. Springer, Heidelberg (2010)
11. Popuri, K., Cobzas, D., Murtha, A., Jägersand, M.: 3D variational brain tumor segmentation using dirichlet priors on a clustered feature set. Int. J. CARS (2011)
12. Prastawa, M., Bullitt, E., Ho, S., Gerig, G.: A brain tumor segmentation framework based on outlier detection. Medical Image Analysis (2004)
13. Schmidt, M., Levner, I., Greiner, R., Murtha, A., Bistriz, A.: Segmenting brain tumors using alignment-based features. In: Proc. of ICMLA (2005)
14. Smith, S.M.: Fast robust automated brain extraction. Hum. Br. Map. (2002)
15. Verma, R., Zacharaki, E.I., Ou, Y., Cai, H., Chawla, S., Lee, A.-K., Melhem, E.R., Wolf, R., Davatzikos, C.: Multi-parametric tissue characterisation of brain neoplasm and their recurrence using pattern classification of MR images. Acad. Radiol. 15(8) (2008)
16. 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)
17. Wen, P.Y., Macdonald, D.R., Reardon, D.A., Cloughesy, T.F., Sorensen, A.G., Galanis, E., Degroot, J., Wick, W., Gilbert, M.R., Lassman, A.B., Tsien, C., Mikkelsen, T., Wong, E.T., Chamberlain, M.C., Stupp, R., Lamborn, K.R., Vogelbaum, M.A., van den Bent, M.J., Chang, S.M.: Updated response assessment criteria for high-grade gliomas: response assessment in neuro-oncology working group. Am. J. Neuroradiol. (2010)

Efficient Global Optimization Based 3D Carotid AB-LIB MRI Segmentation by Simultaneously Evolving Coupled Surfaces

Eranga Ukwatta[1,2], Jing Yuan[1], Martin Rajchl[1,2], and Aaron Fenster[1,2]

[1] Robarts Research Institute
[2] Biomedical Engineering Graduate Program, The University of Western Ontario,
London, ON, Canada
{eukwatta,mrajchl,afenster}@robarts.ca, cn.yuanjing@gmail.com

Abstract. Magnetic resonance (MR) imaging of carotid atherosclerosis biomarkers are increasingly being investigated for the risk assessment of vulnerable plaques. A fast and robust 3D segmentation of the carotid adventitia (AB) and lumen-intima (LIB) boundaries can greatly alleviate the measurement burden of generating quantitative imaging biomarkers in clinical research. In this paper, we propose a novel global optimization-based approach to segment the carotid AB and LIB from 3D T1-weighted black blood MR images, by simultaneously evolving two coupled surfaces with enforcement of anatomical consistency of the AB and LIB. We show that the evolution of two surfaces at each discrete time-frame can be optimized exactly and globally by means of convex relaxation. Our continuous max-flow based algorithm is implemented in GPUs to achieve high computational performance. The experiment results from 16 carotid MR images show that the algorithm obtained high agreement with manual segmentations and achieved high repeatability in segmentation.

Keywords: Carotid atherosclerosis, convex relaxation, continuous max-flow, image segmentation, GPGPU, coupled level sets.

1 Introduction

Stroke is the second leading cause of death worldwide and approximately 87% of the stroke cases are ischemic [1]. Atherosclerosis at the carotid bifurcation is a major cause of generation of thrombosis and subsequent cerebral emboli. Non-invasive, imaging-based biomarkers provide a direct measurement of plaque burden for monitoring plaque progression and regression in patients who undergo medical interventions [2]. MR imaging has shown promise in quantifying carotid measurements, such as vessel wall volume and thickness maps, plaque composition, and inflammation [2], for assessing the efficacy of medical treatment. A robust 3D segmentation of the carotid adventitia (AB) and lumen-intima (LIB) boundaries would greatly assist a comprehensive analysis of carotid atherosclerosis aiding in the translation of these measurements to clinical research.
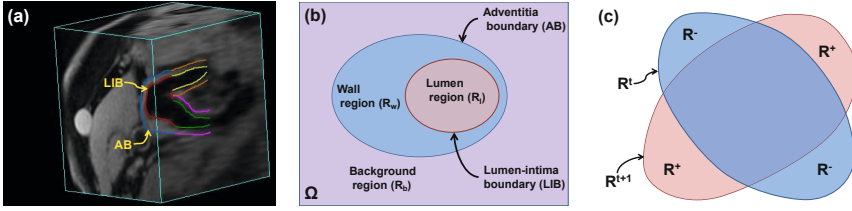
Fig. 1. (a) Carotid MR image with overlaid manual segmentations; (b) Pictorial representation of image domain Ω and its sub-domains R_l, R_w, and R_b; and (c) Contour evolution

Most existing techniques for the carotid AB and LIB MR segmentation are 2D slice-by-slice based methods and delineate the two boundaries separately and independently: Klooster *et al.* [3] used surface fitting for segmentation of the carotid LIB and AB; Kerwin *et al.* [4] proposed to utilize B-spline snakes to extract both boundaries; Adame *et al.* [5] applied fuzzy clustering for carotid LIB segmentation and ellipse fitting with dynamic programming for carotid AB segmentation. However, such methods provide only locally optimal results which are often inefficient and sensitive to the initialization; in addition, they need to explicitly handle changes in topology at the carotid bifurcations and require a great amount of user interaction.

Contributions. In this paper, we describe a novel approach to segment the AB and LIB of common carotid artery (CCA), internal carotid artery (ICA), and external carotid artery (ECA) from T1-weighted (T1w) black blood MR images efficiently and robustly. The segmentation of the carotid AB and LIB is achieved by simultaneously evolving two surfaces with the enforcement of their anatomical order. We show the simultaneous evolution of the coupled surfaces can be solved globally and exactly at each discrete time step, by means of convex relaxation. We propose a continuous max-flow model for the coupled surface evolution, which introduces a dual model to the convex relaxed formulation and derives a fast and fully parallelized algorithm. The results of the experiments demonstrate that our method provides high accuracy and repeatability with significantly less user interaction. The method is mainly intended for monitoring atherosclerosis in patients during medical treatment.

2   Method

The segmentation task of the carotid AB and LIB partitions an MR image into 3 regions: the lumen $\mathcal{R}_l$, the outer-wall $\mathcal{R}_w$, and the background $\mathcal{R}_b$ [see Fig. 1 (a) & (b)]. Since the lumen region is always enclosed within the outer wall region, we model such spatial consistency by a geometrical constraint such that

$$\mathcal{R}_l \subset \mathcal{R}_w\,, \tag{1}$$

where the background region $\mathcal{R}_b = \Omega \backslash \mathcal{R}_w$: i.e. only two regions $\mathcal{R}_l$ and $\mathcal{R}_w$ need to be segmented. Here, we propose a novel global optimization-based approach

for the simultaneous evolution of two surfaces ($\mathcal{R}_l$ and $\mathcal{R}_w$ in this case) subject to histogram-matching and the overlap constraint (1).

2.1   Optimization Model

PDF Matching. The probability density functions (PDFs) of intensities are regional descriptors of the objects of interest, thus matching PDFs provides a robust partitioning of regions with similar PDFs [6, 7]. In practice, model PDFs (i.e. histogram) can be obtained either from a training set or from sampled voxels. In this work, we generated the model PDFs using sampled voxels.

Let $I(x) \in \mathcal{Z}$ be a given 3D carotid MR image, where $\mathcal{Z}$ is the set of image intensity values. $u_i(x)$, $i = l, w, b$, be the indicator function of the estimated region $\mathcal{R}_i$ such that

$$u_i(x) := \begin{cases} 1 \text{, where } x \text{ is inside } \mathcal{R}_i \\ 0 \text{, otherwise} \end{cases}, \quad i = l, w, b. \tag{2}$$

Given that $\mathcal{R}_b = \Omega \backslash \mathcal{R}_w$, we have $u_b = 1 - u_w$, i.e. only two indicator functions $u_l(x)$ and $u_w(x)$ are used in our optimization problem. The PDF $p_i(u, z)$, where $z \in \mathcal{Z}$ and $i = l, w, b$, of the estimated region $\mathcal{R}_i$ is computed using the Parzen method such that

$$p_i(u, z) = \frac{\int_\Omega K(z - I(x))\, u \, dx}{\int u \, dx}, \quad i = l, w, b,$$

where $K(\cdot)$ is the Gaussian kernel function $K(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-x^2/2\sigma^2)$.

Let $q_i(z)$, $i = l, w, b$, be the intensity PDF of region $\mathcal{R}_i$, where $z \in \mathcal{Z}$. We use the statistical divergence metric, e.g. Bhattacharyya or Kullback-Leibler distance etc., to measure the distance between the estimated PDFs $p_i(u, z)$, $i = l, w, b$, of the three regions and their respective model PDFs $q_i(z)$. In this work, the Bhattacharyya distance [7] is used for PDF matching:

$$E_{\text{match}}(u) = -\sum_{i=l,w,b} \sum_{z \in \mathcal{Z}} \sqrt{p_i(u, z)\, q_i(z)}. \tag{3}$$

Optimization Formulation. Using the binary indicator functions $u_{l,w}(x)$, the geometrical overlap prior (1) for regions $\mathcal{R}_l$ and $\mathcal{R}_w$ reduces to the linear inequality constraint

$$u_w(x) \geq u_l(x), \quad \forall x \in \Omega, \tag{4}$$

which enforces the label order of $u_l(x)$ and $u_w(x)$.

In view of the histogram matching energy function (3) and the geometrical overlap constraint (4), we propose to segment 3D carotid MR images by minimizing the energy functional

$$\min_{u_{l,w}(x) \in \{0,1\}} E_{\text{match}}(u) + \sum_{i=l,w} \int_\Omega g_i(x)\, |\nabla u_i(x)| \, dx, \quad \text{s.t. } u_w(x) \geq u_l(x), \tag{5}$$

where $g_i(x) = \lambda_1 + \lambda_2 \times \exp(-\lambda_3 \times |\nabla I(x)|)$ and λ_1, λ_2, and λ_3 are positive constants and $|\nabla I(x)|$ is the normalized image gradient. The anisotropic total-variation term encodes the segmentation with the minimal geodesic length or area. In this work, we propose to minimize the energy functional (5) by means of contour evolution.

2.2   Evolution of Coupled Contours

Distinct from the conventional local optimization-based contour evolution methods, e.g. [7–9] etc, we simultaneously propagate two coupled contours by means of global optimization. The global optimization approaches to evolve a single contour was previously studied by using graph-cuts [10] and by a total variation-based method [11]. In this work, we extend the single contour evolution [10] to evolve two coupled surfaces simultaneously in a spatially continuous setting.

For each region $\mathcal{R}^t$ at time t, we consider its changes w.r.t. its new position $\mathcal{R}$ at the next time frame $t + 1$, in terms of two distinct regions $\mathcal{R}^+$ and $\mathcal{R}^-$ with their respective cost functions $e^+(x)$ and $e^-(x)$ (see Fig. 1)

1. $\mathcal{R}^+$ denotes the increased area w.r.t. $\mathcal{R}$, i.e. for $\forall x \in \mathcal{R}^+$, it is initially outside $\mathcal{R}^t$, but 'jumps' to be inside $\mathcal{R}$; for such 'jump', it pays the cost $e^+(x)$.
2. $\mathcal{R}^-$ denotes the reduced area w.r.t. $\mathcal{R}^t$: for $\forall x \in \mathcal{R}^-$, it is initially inside $\mathcal{R}^t$, but 'jumps' to be outside $\mathcal{R}$ at $t + 1$; for such 'jump', it pays the cost $e^-(x)$.

$\mathcal{R}^t$ is propagated to $\mathcal{R}$ by optimizing the following energy globally and exactly:

$$\min_{\mathcal{R}} \int_{\mathcal{R}^+} e^+(x)\, dx + \int_{\mathcal{R}^-} e^-(x)\, dx + \int_{\partial \mathcal{R}} g(s)\, ds\,. \tag{6}$$

In view of (6), we propose to minimize (5) by evolving the current two regions $\mathcal{R}_l^t$ and $\mathcal{R}_w^t$ to their new positions $\mathcal{R}_l$ and $\mathcal{R}_w$ such that

$$\min_{\mathcal{R}_l, \mathcal{R}_w} \sum_{i=l,w} \left\{ \int_{\mathcal{R}_i^+} e_i^+(x)\, dx + \int_{\mathcal{R}_i^-} e_i^-(x)\, dx \right\} + \sum_{i=l,w} \int_{\partial \mathcal{R}_i} g_i(s)\, ds \tag{7}$$

subject to the geometrical overlap constraint $\mathcal{R}_l \subset \mathcal{R}_w$.

By using the indicator functions (2), we can equally reformulate (7) as follows

$$\min_{u_{l,w} \in \{0,1\}} \langle u_l, C_l \rangle + \langle u_w - u_l, C_w \rangle + \langle 1 - u_w, C_b \rangle + \sum_{i=l,w} \int_{\Omega} g_i(x) |\nabla u_i|\, dx \tag{8}$$

subject to the labeling order constraint (4), i.e. $u_w(x) \geq u_l(x)$ for $\forall x \in \Omega$. The above cost functions $C_i(x)$, $i = l, w, b$, are defined as the first-order variation of the histogram matching term (3) w.r.t. u_l and u_w (see [7] for details).

2.3   Convex Relaxation and Continuous Max-Flow Approach

It has been proven that the non-convex optimization problem (8) can be solved globally and exactly by its convex relaxation [12]:

$$\min_{u_{l,w} \in [0,1]} \langle u_l, C_l \rangle + \langle u_w - u_l, C_w \rangle + \langle 1 - u_w, C_b \rangle + \sum_{i=l,w} \int_{\Omega} g_i(x) |\nabla u_i|\, dx \tag{9}$$

(a) Initialization   (b) Final result   (c) Initialization   (d) Final result
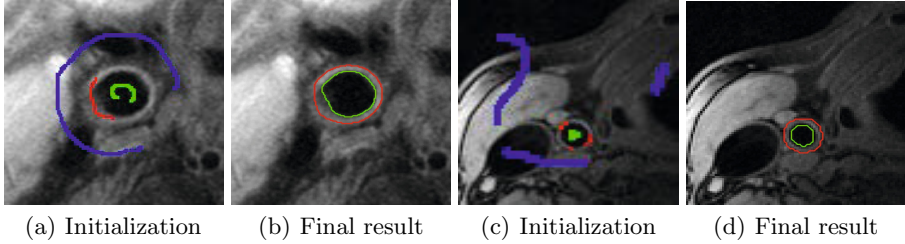
Fig. 2. Example expert initializations and 2D segmentation results. The only user interaction used in the pipeline is the choice of sample seeds on a single transverse slice. Green, red, and blue seeds correspond to lumen, wall and background regions respectively. (a) & (b): T1w 3T MR image.(c) & (d) T1w 1.5T MR image.

subject to the label ordering constraint $u_w(x) \geq u_l(x)$ for $\forall x \in \Omega$. In other words, the two regions $\mathcal{R}_l^t$ and $\mathcal{R}_w^t$ can be evolved to their globally optimal positions $\mathcal{R}_l$ and $\mathcal{R}_w$ at each time frame from t to $t+1$, subject to the geometrical overlap constraint (1). In this work, we follow the continuous min-cut/max-flow theory proposed by Yuan *et al.* [13] and Bae *et al.* [12] to solve the proposed optimization problem (9), globally and exactly. To this end, we adopt their flow configuration [12] and propose the continuous max-flow as follows:

$$\max_{p_b, p_l, p_w} \int_\Omega p_l(x)\, dx \tag{10}$$

subject to the flow capacities

$$|q_i(x)| \leq g_i(x), \quad i = l, w; \quad p_i(x) \leq C_i(x), \quad i = b, l, w; \tag{11}$$

and the flow conservation conditions

$$(\operatorname{div} q_l - p_w + p_l)(x) = 0, \quad (\operatorname{div} q_w - p_b + p_w)(x) = 0, \quad \forall x \in \Omega. \tag{12}$$

The continuous max-flow model is dual/equivalent to the convex relaxation problem (9) [12]. By (10), we derive an efficient continuous max-flow based algorithm which is different from the ones proposed by Bae *et al.* [12]. Our model explores the optimization over all the dual flow functions in parallel, instead of sequentially or group-wise sequentially. In practice, the new parallelized scheme achieves a faster convergence.

3   Experiments and Results

Segmentation Pipeline. Initially, the anisotropic voxels are interpolated into isotropic voxels. In our approach, the carotid wall, lumen and the background regions were initialized using an interactive user interface by the expert *only on a single transverse slice* as shown in Fig. 2 (a) and (c). The sampled voxels are used to generate model PDFs for histogram matching [see (3)] and are considered

<div align="center">

(a) Initial for LIB   (b) Final result   (c) Initial for LIB   (d) Final result

</div>
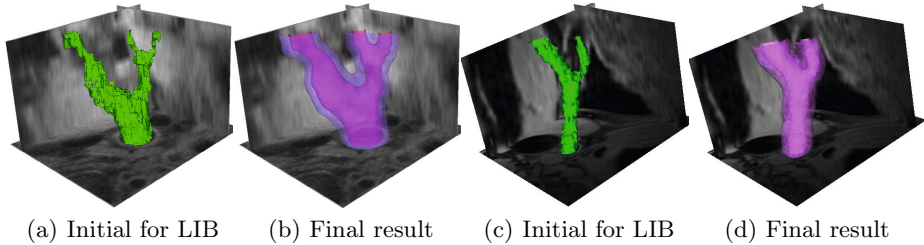
Fig. 3. Example segmentations of a T1w 3T MR image [Fig. 3(b)] and a 1.5T MR image [Fig. 3(d)] by the proposed coupled surface evolution approach

as region-based hard constraints. The carotid AB and LIB are segmented using 2D multi-surface evolution by minimizing the objective function (5). The 2D segmentation result is then used to generate a more representative prior PDFs for each region for the 3D coupled segmentation. We perform a region growing segmentation of the LIB to obtain a crude initial guess [see Fig. 3 (a) & (c)]. We used region growing method for its simplicity; however, the operator could also provide additional sampled seeds on the long-axis direction of the artery. We obtain an initial estimate of the AB surface by dilating the LIB initial surface. Finally, we use the coupled surface evolution in 3D for the segmentation of carotid AB and LIB by minimizing the objective function (5). To incorporate an anatomically-motivated separation of the carotid AB and LIB, we assigned $e^+(x)$ infinite cost for the voxels that are within a minimum separation distance (0.6 mm is used for experiments) outside to the LIB.

Data and Validation. The data comprise of 16 left and right carotid artery T1w black blood MR images from eight subjects: eight 3T (voxel size $\approx 0.2 \times 0.2 \times 2 \text{ mm}^3$) and eight 1.5T (voxel size $\approx 0.5 \times 0.5 \times 2 \text{ mm}^3$) MR images. Subjects were scanned using a GE Excite HD MRI (Milwaukee, WI, USA) with a custom-built six-element carotid-bifurcation-optimized receive-only phased-array coil. The imaging parameters are as follows: TR is 1RR and TEs are 11.4 ms and 12 ms for 3T and 1.5T images respectively with FSE and fat saturation.

The performance of the algorithm was evaluated with respect to manual segmentations in terms of accuracy and reproducibility. Manual segmentations were performed on a slice-by-slice basis on transverse view using a multi-planar reformatting software with 1 mm inter-slice distance up to 4 cm along carotid including CCA, ICA, and ECA. We used Dice coefficient (DSC) as the region-based metric, root mean square error (RMSE), and Hausdorff distance (MAXD) as distance-based metrics.

Results. The convex max-flow algorithm was implemented using parallel computing architecture (CUDA, NVIDIA Corp., Santa Clara, CA) and the user interface in Matlab (Natick, MA). The experiments were conducted on a Quad core Windows workstation with 2.8 GHz and a GPU of NVDIA GTX580. Our algorithm required $\approx 25\text{s}$ of time for expert initialization. The computational

Table 1. Results for eight 3T MR images and eight 1.5T MR images

	Metric	CCA		ICA		ECA	
		AB	LIB	AB	LIB	AB	LIB
3T	DSC (%)	94 ± 2	94 ± 4	87 ± 6	90 ± 5	80 ± 6	83 ± 6
	RMSE (mm)	0.3 ± 0.1	0.4 ± 0.3	0.3 ± 0.1	0.3 ± 0.1	0.3 ± 0.1	0.3 ± 0.2
	MAXD (mm)	1.2 ± 0.2	0.5 ± 0.6	2 ± 1.6	0.7 ± 0.2	1.4 ± 1.5	0.8 ± 0.3
1.5T	DSC (%)	90 ± 5	93 ± 3	75 ± 9	92 ± 4	69 ± 12	88 ± 4
	RMSE (mm)	0.5 ± 0.2	0.8 ± 1.2	0.6 ± 0.3	0.4 ± 0.1	0.6 ± 0.3	0.4 ± 0.1
	MAXD (mm)	1.6 ± 1.0	1.5 ± 2.0	3.1 ± 5.3	1.1 ± 1.1	6 ± 4.0	0.9 ± 0.3

Table 2. Results for observer variability using DSC(%) for 8 1.5T MR images

Repetition #	CCA		ICA		ECA	
	AB	LIB	AB	LIB	AB	LIB
1	89.8 ± 5.1	93.3 ± 3.2	75.5 ± 9.0	91.8 ± 3.9	70.4 ± 12.0	87.9 ± 3.8
2	89.6 ± 5.0	93.3 ± 3.3	75 ± 8.9	92 ± 4.0	68.2 ± 11.5	88.4 ± 4.5
3	89.7 ± 5.0	93.6 ± 3.4	75.2 ± 9.1	92.2 ± 4.3	69.3 ± 10.7	88.3 ± 4.1

time for convergence of the algorithm was ≈ 40s (8s for max-flow in a GPU and 32s for cost computation using non-optimized Matlab code) which was achieved within 10-15 iterations for a single 3D MR image with 110 slices. Figure 3 (b) & (d) show the carotid AB and LIB surfaces generated using our algorithm for some example 3T and 1.5T MR images. The performance results of the algorithm are shown in Table 1. The algorithm yielded high DSC and low RMSE for CCA, ICA, and ECA for both data sets except for the ECA AB where a DSC of 70% was obtained for 1.5T MR images. We also assessed the intra-observer variability by repeatedly segmenting the same image set three times with different initializations. The results of the intra-observer variability analysis of the algorithm is shown in Table 2. The algorithm yielded approximately similar DSCs in all three repetitions, which suggests a high reproducibility of our approach.

4   Discussion and Conclusion

We developed a novel global optimization approach for coupled surface evolution for carotid AB and LIB segmentations. The coupling permits the integration of image information derived from both surfaces to drive their optimization. The algorithm provided robust and efficient segmentation results for the AB and LIB in terms of accuracy and intra-observer reproducibility. The 2D method proposed by Adame *et al.* [5] is currently used in clinical trials, which requires 40s to segment a single 2D slice. Our method provides substantial improvement in speed for segmenting 3D images over previous methods. Li *et al.* [14] also proposed a single-shot graph-cut approach to segment coupled surfaces, but their method need to unwrap the image domain, in which handling carotid bifurcations or changes in topology are challenging.

Most previous papers segment only the carotid CCA and report accuracy in terms of area measurements but not DSC or RMSE [5, 4]. Our algorithm reports a higher accuracy in terms of DSC for the CCA and ICA than ECA. The reduced accuracy for the ECA is mostly due to the weak image information of the ECA AB, in which the algorithm attempts to maintain a minimal length surface without splitting at the carotid bifurcations. Clinically, CCA and ICA segmentations are more important than ECA as the plaque tend to be present mostly in CCA and ICA [2]. The expert may initialize on more than one 2D slice to achieve further accuracy of the segmentations, especially proximal to the carotid bifurcations. The performance results of our algorithm suggest that it may be suitable for use in clinical trials involving the monitoring of carotid atherosclerosis using 3D MR imaging-based biomarkers.

References

1. Roger, V., Go, A., Lloyd-Jones, D., Adams, R., Berry, J., Brown, T., et al.: Heart disease and stroke statistics 2011 update1. Circulation 123(4), e18–e209 (2011)
2. Yuan, C., Oikawa, M., Miller, Z., Hatsukami, T.: MRI of carotid atherosclerosis. J. Nucl. Cardiol. 15(2), 266–275 (2008)
3. van't Klooster, R., de Koning, P.J., Dehnavi, R.A., Tamsma, J.T., de Roos, A., Reiber, J.H., van der Geest, R.J.: Automatic lumen and outer wall segmentation of the carotid artery using deformable 3D models in MR angiography and vessel wall images. JMRI 35(1), 156–165 (2011)
4. Kerwin, W., Xu, D., Liu, F., Saam, T., Underhill, H., Takaya, N., Chu, B., Hatsukami, T., Yuan, C.: Magnetic resonance imaging of carotid atherosclerosis. Top. Magn. Reson. Imag. 18(5), 371–378 (2007)
5. Adame, I.M., van der Geest, R.J., Wasserman, B.A., Mohamed, M.A., et al.: Automatic segmentation and plaque characterization in atherosclerotic carotid artery MR images. Magn. Reson. Mater. Phy. 16(5), 227–234 (2004)
6. Aubert, G., Barlaud, M., Faugeras, O., Jehan-Besson, S.: Image segmentation using active contours: calculus of variations or shape gradients? SIAM J. Appl. Math. 63(6), 2128–2154 (2003)
7. Michailovich, O., Rathi, Y., Tannenbaum, A.: Image segmentation using active contours driven by the Bhattacharyya gradient flow. IEEE TIP 16(11), 2787–2801 (2007)
8. Caselles, V., Kimmel, R., Sapiro, G.: Geodesic active contours. IJCV 22(1) (1997)
9. Chan, T., Vese, L.A.: Active contours without edges. IEEE TIP 10, 266–277 (2001)
10. Boykov, Y., Kolmogorov, V., Cremers, D., Delong, A.: An Integral Solution to Surface Evolution PDEs Via Geo-cuts. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3953, pp. 409–422. Springer, Heidelberg (2006)
11. Chambolle, A.: An algorithm for mean curvature motion. Interf. Free Bound. 6, 195–218 (2004)
12. Bae, E., Yuan, J., Tai, X.C., Boycov, Y.: A fast continuous max-flow approach to non-convex multilabeling problems. Technical report CAM-10-62, UCLA (2010)
13. Yuan, J., Bae, E., Tai, X.: A study on continuous max-flow and min-cut approaches. In: CVPR 2010 (2010)
14. Li, K., Wu, X., Chen, D., Sonka, M.: Optimal surface segmentation in volumetric images-a graph-theoretic approach. IEEE T. Pattern. Anal. 28(1), 119–134 (2006)

Sparse Patch Based Prostate Segmentation in CT Images

Shu Liao[1], Yaozong Gao[1,2], and Dinggang Shen[1]

[1] Department of Radiology and BRIC, University of North Carolina at Chapel Hill
liaoshu.cse@gmail.com
[2] Department of Computer Science, University of North Carolina at Chapel Hill
yzgao@cs.unc.edu, dgshen@med.unc.edu

Abstract. Automatic prostate segmentation plays an important role in image guided radiation therapy. However, accurate prostate segmentation in CT images remains as a challenging problem mainly due to three issues: Low image contrast, large prostate motions, and image appearance variations caused by bowel gas. In this paper, a new patient-specific prostate segmentation method is proposed to address these three issues. The main contributions of our method lie in the following aspects: (1) A new patch based representation is designed in the discriminative feature space to effectively distinguish voxels belonging to the prostate and non-prostate regions. (2) The new patch based representation is integrated with a new sparse label propagation framework to segment the prostate, where candidate voxels with low patch similarity can be effectively removed based on sparse representation. (3) An online update mechanism is adopted to capture more patient-specific information from treatment images scanned in previous treatment days. The proposed method has been extensively evaluated on a prostate CT image dataset consisting of 24 patients with 330 images in total. It is also compared with several state-of-the-art prostate segmentation approaches, and experimental results demonstrate that our proposed method can achieve higher segmentation accuracy than other methods under comparison.

1 Introduction

Prostate cancer is the second leading cause of cancer death for male in US. Image guided radiation therapy (IGRT), as a non-invasive approach, is one of the major treatment methods for prostate cancer. The key to the success of IGRT is the accurate localization of prostate in the treatment images such that cancer cells can be effectively eliminated by the high energy X-rays delivered to the prostate.

However, accurate prostate localization in CT images remains as a challenging problem. First, the image contrast between the prostate and its surrounding tissues is low. This issue can be illustrated by Figures 1 (a) and (b). Second, the prostate motion across different treatment days can be large. Third, the image appearance can be significantly different due to the uncertain existence of bowel gas. This issue can be illustrated by Figures 1 (a) and (c).
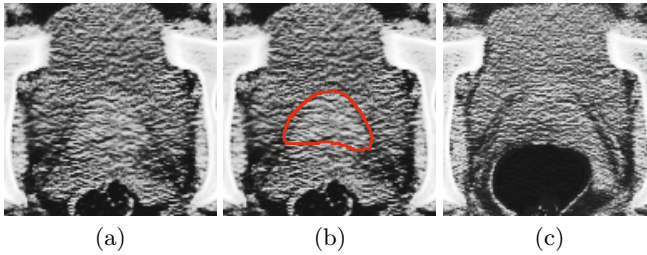
Fig. 1. (a) An image slice of a 3D prostate CT image volume. (b) The manually delineated prostate boundary (i.e., the red contour) by the radiologist superimposed on the image in (a). (c) An image slice with bowel gas of a 3D prostate CT image volume with the same patient as in (a) but acquired from a different treatment day.

Many novel approaches have been proposed in the literature [1–4] for prostate segmentation in CT images. For instance, Davis *et al.* [1] proposed a deflation algorithm to eliminate the distortions brought by the existence of bowel gas, followed by a diffeomorphic image registration process to localize the prostate. Chen *et al.* [3] proposed a Bayesian framework for prostate segmentation in CT images. In this paper, we propose a new patient-specific prostate segmentation method. The main contributions of our method lie in the following aspects: (1) Anatomical features are extracted from each voxel position. The most informative features are selected by the logistic Lasso to form the new salient patch based signature. (2) The new patch based signature is integrated with a sparse label propagation framework to localize the prostate in each treatment image. (3) An online update mechanism is adopted to capture more patient-specific information from the segmented treatment images. Our method has been extensively evaluated on a prostate CT image dataset consisting of 24 patients with 330 images in total. It is also compared with several state-of-the-art prostate segmentation approaches, and experimental results demonstrate that our method achieves higher segmentation accuracy than other methods under comparison.

2   Patch Based Signature in Discriminative Feature Space

Patch based representation has been widely used in medical image analysis [5] as anatomical signatures for each voxel. The conventional patch based principle is to define a small $K \times K$ image patch centered at each voxel $\boldsymbol{x}$ as the signature of $\boldsymbol{x}$, where K denotes the scale of interest.

However, the conventional patch based signature still may not fully distinguish voxels belonging to the prostate and non-prostate regions due to the low image contrast around the prostate boundary. Therefore, we are motivated to construct more salient patch based signatures in the feature space. For each image $I(\boldsymbol{x})$, it is convolved with a specific kernel function $\psi_j(\boldsymbol{x})$ by Equation 1:

$$F_j(\boldsymbol{x}) = I(\boldsymbol{x}) * \psi_j(\boldsymbol{x}), \quad (j = 1, ..., L) \tag{1}$$

where $F_j(\boldsymbol{x})$ denotes the resulting feature value with respect to the jth kernel $\psi_j(\boldsymbol{x})$ at voxel $\boldsymbol{x}$. L denotes the number of kernel functions used to extract features. In this paper, we adopt the Haar [6], histogram of oriented gradient (HOG) [7], and local binary pattern (LBP) [8] features.

For each feature map $F_j(\boldsymbol{x})$ (j=1,...,L), we can also define a $K \times K$ patch centered at each voxel $\boldsymbol{x}$. Thus, $\boldsymbol{x}$ has a $K \times K \times L$ dimensional signature, denoted as $\boldsymbol{a}(\boldsymbol{x})$. Then, the logistic Lasso [9] is adopted to select the most informative features. More specifically, N voxels $\boldsymbol{x}_i$ (i=1,...,N) are sampled from the training images, with labels $l_i = 1$ if $\boldsymbol{x}_i$ belongs to the prostate and $l_i = -1$ otherwise. Then, we aim to minimize the logistic Lasso problem [10] in Equation 2:

$$J(\boldsymbol{\beta}, c) = \sum_{i=1}^{N} \log(1 + \exp(-l_i(\boldsymbol{\beta}^T \boldsymbol{a}(\boldsymbol{x}_i) + c))) + \lambda||\boldsymbol{\beta}||_1, \tag{2}$$

where $\boldsymbol{\beta}$ is the sparse coefficient vector, $||\cdot||_1$ is the L1 norm, c is the intercept scalar, and λ is the regularization parameter. The optimal solution $\boldsymbol{\beta}^{opt}$ and c^{opt} to minimize Equation 2 can be estimated by Nesterov's method [11]. The most informative features can be determined by selecting features with $\boldsymbol{\beta}^{opt}(d) \neq 0$ ($d = 1, ..., K \times K \times L$). We denote the final signature of each voxel $\boldsymbol{x}$ as $\boldsymbol{b}(\boldsymbol{x})$.

3   Sparse Patch Based Label Propagation

The general label propagation process [5] can be formulated as follows: Given n training images and their segmentation groundtruths (i.e., label maps), denoted as $\{(I_u, S_u), u = 1, ..., n\}$. For a new treatment image I_{new}, each voxel $\boldsymbol{x}$ in I_{new} is linked to each voxel $\boldsymbol{y}$ in I_u with a graph weight $w_u(\boldsymbol{x}, \boldsymbol{y})$. Then, the corresponding label for each voxel $\boldsymbol{x}$ can be estimated by Equation 3:

$$S_{new}(\boldsymbol{x}) = \frac{\sum_{u=1}^{n} \sum_{\boldsymbol{y} \in \Omega} w_u(\boldsymbol{x}, \boldsymbol{y}) S_u(\boldsymbol{y})}{\sum_{u=1}^{n} \sum_{\boldsymbol{y} \in \Omega} w_u(\boldsymbol{x}, \boldsymbol{y})}, \tag{3}$$

where Ω denotes the image domain, and S_{new} denotes the prostate probability map of I_{new}.

Based on the new patch based signature derived in Section 2, the graph weight $w_u(\boldsymbol{x}, \boldsymbol{y})$ can be defined by Equation 4 similar to [5]:

$$w_u(\boldsymbol{x}, \boldsymbol{y}) = \begin{cases} \Phi\left(\frac{||\boldsymbol{b}_{I_{new}}(\boldsymbol{x}) - \boldsymbol{b}_{I_u}(\boldsymbol{y})||_2}{2K\alpha}\right), & \text{if } \boldsymbol{y} \in \mathcal{N}_u(\boldsymbol{x}); \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

where $\boldsymbol{b}_{I_{new}}(\boldsymbol{x})$ and $\boldsymbol{b}_{I_u}(\boldsymbol{y})$ denote the patch signature at $\boldsymbol{x}$ of I_{new} and at $\boldsymbol{y}$ of I_u, respectively, and $||\cdot||_2$ denotes the L2 norm. The heat kernel $\Phi(x) = e^{-x}$ is used similar to [5]. α is the smoothing parameter. $\mathcal{N}_u(\boldsymbol{x})$ denotes the neighborhood of voxel $\boldsymbol{x}$ in image I_u, and it is defined as the $W \times W \times W$ subvolume centered at $\boldsymbol{x}$. Noted that the standard deviation of image noise for the selected features could also be considered in Equation 4 to estimate the graph weight more accurately.
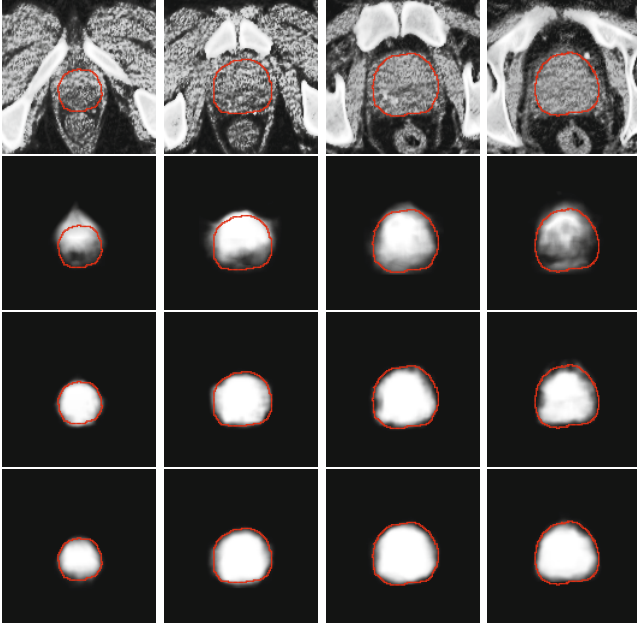
Fig. 2. First row: original prostate CT images, overlayed with the groundtruth highlighted by red contours. Second to fourth rows: The corresponding probability maps obtained by label propagation with the intensity patch voxel signature, patch based voxel signature in the feature space without using sparse representation, and the patch based voxel signature in the feature space with sparse representation, respectively

Inspired by the discriminant power of sparse representation, we are motivated to estimate the graph weight based on Lasso. Specifically, we organize $\boldsymbol{b}_{I_u}(\boldsymbol{y})$, $\boldsymbol{y} \in \mathcal{N}_u(\boldsymbol{x})$ $(u = 1, ..., n)$ as columns of a matrix $\mathbf{A}$. Then, the sparse coefficient vector $\boldsymbol{\theta_x}$ for voxel $\boldsymbol{x}$ is estimated by minimizing Equation 5 with Nesterov's method [11]:

$$J(\boldsymbol{\theta_x}) = \frac{1}{2}||\boldsymbol{b}_{I_{new}}(\boldsymbol{x}) - \mathbf{A}\boldsymbol{\theta_x}||_2^2 + \lambda||\boldsymbol{\theta_x}||_1, \quad \boldsymbol{\theta_x} \geq 0. \tag{5}$$

We denote the optimal solution of Equation 5 as $\boldsymbol{\theta_x^{opt}}$. The graph weight $w_u(\boldsymbol{x}, \boldsymbol{y})$ is set to the corresponding element in $\boldsymbol{\theta_x^{opt}}$ with respect to $\boldsymbol{y}$ in image I_u.

Then, the prostate probability map S_{new} is estimated by Equation 3. Figure 2 shows an example of the prostate probability maps obtained by different strategies. It is shown that the new patch based signature with sparse label propagation estimates the prostate boundary more accurately.

After estimating the prostate probability map S_{new}, the prostate in I_{new} can be localized by aligning the segmentation groundtruth of each training image S_u $(u = 1, ..., N)$ to S_{new} and perform majority voting. The affine transformation [12] is used due to the relatively simple shape of the prostate.

4   The Online Update Mechanism

In this paper, an online update mechanism is adopted to incorporate more patient-specific information. More specifically, at the beginning stage of the radiation therapy, only the planning image of the current patient is available to serve as the training image. As more treatment images are collected and segmented during the therapy process, they will also serve as the training images. Therefore, the online update mechanism gradually captures more patient-specific information during the period of radiation therapy. It is worth noting that manual adjustments to the automatic segmentation results are also feasible to correct some poorly segmented images since such adjustments can be completed offline.

5   Experimental Results

Our method was evaluated on a 3D prostate CT dataset consisting of 24 patients with 330 images. Each image has in-plane resolution 512×512, with voxel size $0.98 \times 0.98 \ mm^2$. The voxel size along the sagittal direction is $3 \ mm$. The segmentation groundtruth provided by the clinical expert is also available.

In all the experiments, the following parameter settings were adopted for our method by cross validation: patch size $K = 5$, neighborhood size $W = 15$, $\lambda = 10^{-4}$ in Equations 2 and 5, $\alpha = 1$ in Equation 4. Therefore, the dimension of the original feature is 1325, and the same subset of selected feature is used

(a)

(b)

(c)
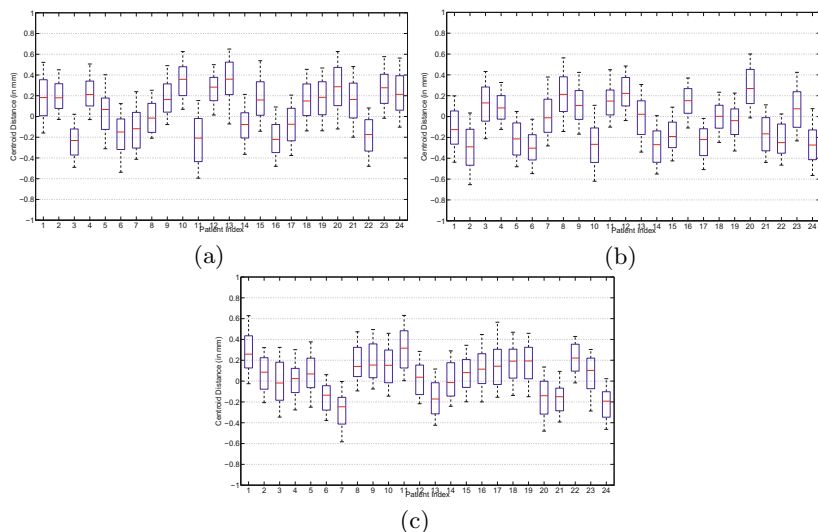
Fig. 3. Centroid distance between the estimated prostate by our method and the groundtruth along the (a) lateral, (b) anterior-posterior, and (c) superior-inferior directions, respectively. The horizontal lines in each box represent the 25th percentile, median, and 75th percentile. The whiskers extend to the most extreme data points.
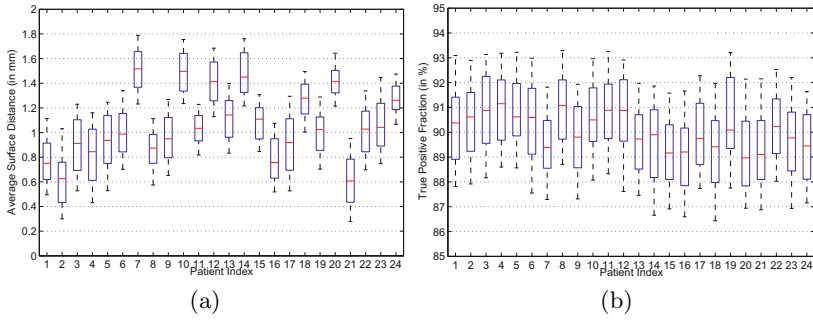
Fig. 4. Whisker plots of the (a) average surface distance and (b) true positive fraction between our estimated prostate and the groundtruth for each patient. The horizontal lines in each box represent the 25th percentile, median, and 75th percentile. The whiskers extend to the most extreme data points.
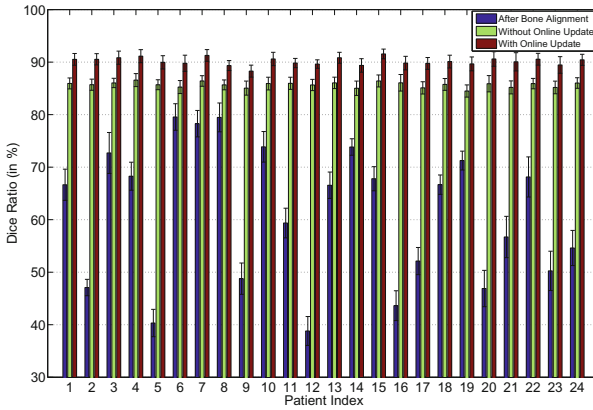
Fig. 5. The average Dice ratios and the standard deviations of the 24 patients after bone alignment, and using our method without or with the online update mechanism.

for all voxels. For each patient, we first rigidly aligned the pelvic bone structures of each treatment image to the planning image with FLIRT [12] to remove the whole-body rigid motion.

Four quantitative measures were used to evaluate the segmentation accuracy of our method: The centroid distance (CD), average surface distance (ASD), true positive fraction (TPF), and the Dice ratio. The whisker plots of CD along the lateral, anterior-posterior and superior-inferior directions of each patient are shown in Figures 3 (a) to (c), respectively. It is shown that the median CD along each direction of each patient lies within 0.4 mm from the groundtruth, which implies the effectiveness of our method. The whisker plots of the ASD and TPF measures of each patient are also shown in Figures 4 (a) to (b), respectively.
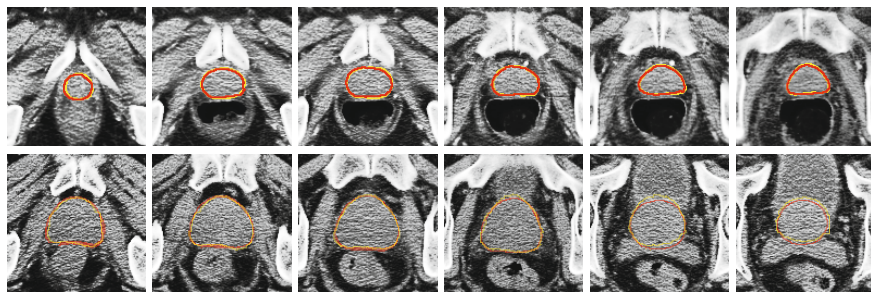
Fig. 6. Typical segmentation results of the 3rd treatment image of the 17th patient (first row, Dice ratio 90.04%) and the 5th treatment image of the 15th patient (second row, Dice ratio 90.86%). The yellow contours denote the estimated prostate boundaries by our method, and the red contours denote the prostate boundaries of the groundtruth.

Table 1. Comparison of different methods, N/A indicates that the corresponding result was not reported in the respective paper. The last three rows denote our method using intensity patch signature, discriminative (Dist) patch signature in feature space, and the discriminative (Dist) patch signature with sparse label propagation. Results obtained by our method on the last row are bolded.

Methods	Mean Dice Ratio	Mean ASD (mm)	Mean CD (x/y/z) (mm)	Median TPF
Davis *et al.* [1]	0.820	N/A	-0.26/0.35/0.22	N/A
Feng *et al.* [2]	0.893	2.08	N/A	N/A
Chen *et al.* [3]	N/A	1.10	N/A	0.84
Intensity Patch	0.731	3.58	1.84/1.62/-1.74	0.72
Dist Patch	0.862	1.52	0.24/-0.23/0.31	0.84
Our Method	**0.909**	**1.09**	**0.17/−0.09/0.19**	**0.89**

The average Dice ratio and its standard deviation between the estimated prostate volume and the groundtruth for each patient after bone alignment, with and without the online update mechanism of our method are also shown in Figure 5. It is shown that our method can achieve high Dice ratios (i.e., mostly above 85%) even without the online update mechanism (i.e, only the planning image is served as the training data). The segmentation accuracies can be further improved with the online update mechanism (i.e., mostly above 90%).

A typical example of the segmentation results by using the proposed method is shown in Figure 6. It can be observed that the estimated prostate boundaries are very closed to the prostate boundaries of the groundtruth even for the apex area of the prostate and in slices where bowel gas exists.

Our method was also compared with three state-of-the-art prostate CT segmentation algorithms [1–3]. The best results reported in [1–3] were adopted for comparison, and detailed comparisons are listed in Table 1. It can be observed from Table 1 that our method outperforms other methods under comparison. The contributions of the new patch based signature and the sparse label propagation framework can also be mirrored by Table 1.

6   Conclusion

In this paper, we propose a new patient-specific prostate segmentation method for CT images. Our method extracts anatomical features from each voxel position, and the most informative features are selected by logistic Lasso to construct the patch based representation in the feature space. It is shown that the new patch based signature can distinguish voxels belonging to the prostate and non-prostate regions more effectively. The new patch based signature is integrated with a sparse label propagation framework to localize the prostate in new treatment images. An online update mechanism is also adopted in this paper to capture the patient-specific information more effectively. The proposed method has been extensively evaluated on a prostate CT image dataset consisting of 24 patients with 330 images. It is also compared with several state-of-the-art prostate segmentation methods. Experimental results show that our method achieves higher segmentation accuracy than other methods under comparison.

References

1. 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. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3749, pp. 442–450. Springer, Heidelberg (2005)
2. Feng, Q., Foskey, M., Chen, W., Shen, D.: Segmenting CT prostate images using population and patient-specific statistics for radiotherapy. Medical Physics 37, 4121–4132 (2010)
3. Chen, S., Lovelock, D., Radke, R.: Segmenting the prostate and rectum in CT imagery using anatomical constraints. Medical Image Analysis 15, 1–11 (2011)
4. Li, W., Liao, S., Feng, Q., Chen, W., Shen, D.: Learning Image Context for Segmentation of Prostate in CT-Guided Radiotherapy. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part III. LNCS, vol. 6893, pp. 570–578. Springer, Heidelberg (2011)
5. Rousseau, F., Habas, P., Studholme, C.: A supervised patch-based approach for human brain labeling. IEEE TMI 30, 1852–1862 (2011)
6. Mallat, G.: A theory for multiresolution signal decomposition: the wavelet representation. IEEE PAMI 11, 674–693 (1989)
7. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: CVPR, pp. 886–893 (2005)
8. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE PAMI 24, 971–987 (2002)
9. Tibshirani, R.: Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B 58, 267–288 (1996)
10. Liu, J., Ji, S., Ye, J.: SLEP: Sparse Learning with Efficient Projections. Arizona State University (2009)
11. Nesterov, Y.: Introductory Lectures on Convex Optimization: A Basic Course. Kluwer Academic Publishers (2004)
12. Jenkinson, M., Bannister, P., Brady, M., Smith, S.: Improved optimization for the robust and accurate linear registration and motion correction of brain images. NeuroImage 17, 825–841 (2002)

Anatomical Landmark Detection Using Nearest Neighbor Matching and Submodular Optimization

David Liu and S. Kevin Zhou

Siemens Corporation, Corporate Research and Technology, Princeton NJ, USA

Abstract. We present a two-stage method for effective and efficient detection of one or multiple anatomical landmarks in an arbitrary 3D volume. The first stage of nearest neighbor matching is to roughly estimate the landmark locations. It searches out of 100,000 volumes for the closest to an input volume and then transfers landmark annotations to the input. The second stage of submodular optimization is to refine the landmark locations by running discriminative landmark detectors within the search ranges constrained by the first stage results. Further it coordinates multiple detectors with a search strategy optimized on the fly to reduce the overall computation cost arising in a submodular formulation. We validate the accuracy, speed and robustness of our approach by detecting body regions and landmarks in a dataset of 2,500 CT scans.

1 Introduction

In the paper, we define an anatomical landmark (or landmark in brief) as a distinct point in a body scan that coincides with anatomical structures, such as liver top, lung top, aortic arch, iliac artery bifurcation, femur head left and right, to name but a few. Landmark detection is crucial for medical image applications. As a body region can be defined by landmark(s)[1], body region detection can be solved by landmark detection. Landmarks also provide seed points to initiate image segmentation [1] and registration[2]. In seminar reporting, the detected organ landmarks can help config the optimal intensity window for display [3] and offer the text tooltips for structures in the scan [4].

A practical landmark detection method must meet the following requirements. First, it must be robust to deal with pathological or anomalous anatomies such as fluid-filled lungs, air-filled colons, inhomogeneous livers caused by different metastasis, and resected livers after surgical interventions, different contrast agent phases, scans of full or partial body regions, extremely narrow field of views, etc. Figure 1 shows some examples of CT scans that illustrate the challenges. Second, since landmark detection is mostly a pre-processing step for computationally heavier tasks such as CAD and registration, it must run fast so that more time can be allocated for the heavier tasks. Finally, the landmark detection accuracy depends on the subsequent applications. For example, for body region detection exact 3D point positions are not needed; for registration, accurate landmarks are desired.

---

[1] A simple approach for determining the body region could rely on certain DICOM tags. But these tags are not always reliable, justifying a need for a dedicated image-based algorithm.
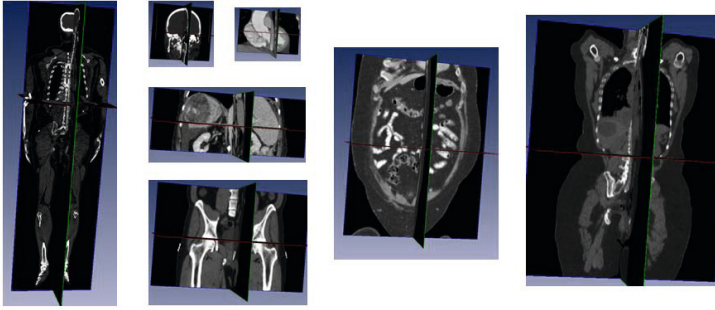
Fig. 1. Our dataset has 2,500 3D CT scans with different body regions and severe pathologies

We leverage the unitary, pairwise, and holistic contextual information (defined in Section 2) manifested in medical images to design an effective and efficient method for detection of one or multiple landmarks in a 3D volume. The proposed algorithm in Section 3 has two stages. First, nearest neighbor (NN) matching is to roughly estimate the landmark locations. It searches out of 100,000 volumes for the NN to the input query volume and then transfers the landmark annotations to the query. Second, submodular optimization is to refine the landmark locations by running discriminative landmark detectors within the search ranges constrained by the results from the first stage. Further it coordinates multiple detectors with a search strategy optimally determined on the fly to further reduce the overall computational cost arising in a submodular formulation. We validate the robustness, speed and accuracy of our approach by detecting one or multiple landmarks in a dataset of 2,500 CT volumes in Section 4.

2   Related Work and Context Exploitation

Designing a useful landmark detection method should effectively exploit the rich contextual information manifested in the body scans, which can be generally categorized as *unitary*, *pairwise or higher-order*, and *holistic* context. The *unitary context* refers to the local regularity surrounding a single landmark. The classical object detection approach in [5,6] exploits the unitary context to learn a series of supervised classifier to separate the positive object (herein landmark) from negative background. The complexity of this approach depends on the volume size. The *pairwise or higher-order context* refers to the joint regularities between two landmarks or among multiple landmarks. Liu et al. [7] embed the pairwise spatial contexts among all landmarks into a submodular formulation that minimizes the combined search range for detecting multiple landmarks. Here the landmark detector is still learned by exploiting the unitary context. In [8], the pairwise spatial context is used to compute the information gain that guides an active scheduling scheme for detecting multiple landmarks. Seifert et al. [9] encoded pairwise spatial contexts into a discriminative anatomical network. The *holistic context* goes beyond the relationships among a cohort of landmarks and refers to the whole relationship between all voxels and the landmarks; in other words, regarding the image as a whole. In [10], shape regression machine is proposed to learn a boosting regression function to

predict the object bounding box from the image appearance bounded in an arbitrarily located box and another regression function to predict the object shape. Pauly et al. [3] simultaneously regress out the locations and sizes of multiple organs with confidence scores using a learned Random Forest regressor. To some extent, image registration [11] can be regarded as using the holistic context too.

The proposed approach leverages all three contexts. The first stage of nearest neighbor (NN) matching exploits the holistic context. But instead of learning a regression function to capture the relationship between all voxels and the landmarks, we directly perform a full match and transfer the landmark annotations. This way we avoid scanning the image. The second stage of submodular optimization builds upon the approach in [7] that exploits both the unitary and pairwise contexts, but minimizes the overall computation instead of the total search range in [7]. Also, no holistic context is used in [7]. Instead, an 'anchor landmark' is first detected before triggering the whole detection process, utilizing only the unitary context in an exhaustive scanning.

3   Landmark Detection

3.1   Stage 1: NN Matching for Coarse Detection

Assume that a volume is represented by a D-dimensional feature vector. Given a query (unseen input) vector $x \in R^D$, the problem is to find the element y^* in a finite set Y of vectors to minimize the distance to the query vector:

$$y^* = \arg\min_{y \in Y} d(x, y) \qquad (1)$$

where $d(.,.)$ is the Euclidean distance function. Other choices can be used too.

Volume Features. To facilitate the matching, we represent each volume by a D-dimensional feature vector. In particular, we adopt a representation of the image using 'global features' that provide a holistic description as in [12], where a 2D image is divided into 4×4 regions, eight oriented Gabor filters are applied over four different scales, and the average filter energy in each region is used as a feature, yielding in total 512 features. For our 3D volumes, we compute such features from nine 2D images, consisting of the sagittal, axial, and coronal planes that pass through 25%, 50%, and 75% of the respective volume dimension, resulting a 4,608-dimensional feature vector.

Efficient NN Search. In practice, finding the closest volume through evaluating the exact distances is too expensive when the database size is large and the data dimensionality is high. Two efficient approximations are used for speedup. Vector quantization is used to address the database size issue and product quantization [13] for the data dimensionality issue.

A quantizer is a function $q(.)$ mapping a D-dimensional vector x to a vector $q(x) \in Q = \{q_i, i = 1, 2, ..., K\}$. The finite set Q is called the codebook, which consists of K centroids. The set of vectors mapped to the same centroid forms a Voronoi cell, defined as $V_i = \{x \in R^D | q(x) = q_i\}$. The K Voronoi cells partition the space of R^D. The quality of a quantizer is often measured by the mean squared error between an input vector and its representative centroid $q(x)$. We use the K-means algorithm to find a

near-optimal codebook. During the search stage, which has high speed requirement, distance evaluation between the query and a database vector consists of computing the distance between the query vector and the nearest centroid of the database vector.

Our volume feature vectors are high dimensional (we use $D = 4608$ dimensions), which poses difficulty for a straightforward implementation of the K-means quantization described above. A quantizer that uses only $1/3$ bits per dimension already has 2^{1536} centroids. Such a large number of centroids makes it impossible to run the K-means algorithm in practice. Product quantization [13] addresses this issue by splitting the high-dimensional feature vector into m distinct sub-vectors as follows,

$$\underbrace{x_1, ..., x_{D^*}}_{u_1(x)}, ..., \underbrace{x_{D-D^*+1}, ..., x_D}_{u_m(x)} \tag{2}$$

The quantization is subsequently performed on the m sub-vectors $q^1(u_1(x)), ..., q^m(u_m(x))$, where $q^i, i = 1, ..., m$ denote m different quantizers. In the special case where $m = D$, product quantization is equivalent to scalar quantization, which has the lowest memory requirement but does not capture any correlation across feature dimensions. In the extreme case where $m = 1$, product quantization is equivalent to traditional quantization, which fully captures the correlation among different features but has the highest (and practically impossible, as explained earlier) memory requirement. We use $m = 1536$ and $K = 4$ (2 bits per quantizer).

Transferring Landmark Annotations. Given a query, we use the aforementioned method to find the most similar database volume. Assume this database volume consists of N landmarks with positions $\{s_1, ..., s_N\}$. We simply 'transfer' these landmark positions to the query. In other words, the coarsely detected landmark positions are set as $\{s_1, ..., s_N\}$. In the next section, we discuss how to refine these positions.

3.2   Stage 2: Submodular Optimization for Refined Detection

After the stage of NN matching, certain landmarks are located roughly. We now trigger the landmark detectors to search for a more precise position for each landmark only within local search ranges predicted from the first stage results. Running a landmark detector locally instead of over the whole volume reduces the computation and also reduces false positive detections. The local search range of each detector is obtained offline based on spatial statistics that capture the relative position of each pair of landmarks. Note that the two sets of landmarks in two stages can be different.

In order to speed up the detection, the order of triggering the landmark detectors needs to be considered. This is because, once a landmark position is refined by a detector, we can further reduce the local search ranges for the other landmarks by using the pairwise spatial statistics that embody the *pairwise context*. Consider a volume with N landmarks. Denote by $\Lambda_{(1):(n)} = \{l_{(1)} \prec l_{(2)} \prec ... \prec l_{(n)}\}, n \leq N$ the ordered set of landmarks that have been refined by detectors. Denote by U the un-ordered set of landmarks that remains to be refined. For each landmark $l_i \in U$, we define its search range $\Omega[l_i|\Lambda_{(1):(n)}]$ as the intersection of the search ranges predicted by the already detected landmarks:

$$\Omega[l_i|\Lambda_{(1):(n)}] = \bigcap_{j, l_j \in \Lambda_{(1):(n)}} \Omega[l_i|\{l_j\}], \tag{3}$$

where $\Omega[l_i|\{l_j\}]$ denotes the local search neighborhood for landmark l_i conditioned on the position of a detected landmark l_j.

Denote the volume of search range $\Omega[l_j|\Lambda]$ as $V(\Omega[l_j|\Lambda])$. Without loss of generality, assume the search volume is the cardinality of the set of voxels that fall within the search range. Denote by $\alpha[l_j]$ the unit computation cost for evaluating the detector for landmark l_j. Our goal is then to find the ordered set $\Lambda_{(1):(N)}$ that minimizes the *total computation*, i.e. ,

$$\Lambda'_{(1):(N)} = \underset{\Lambda_{(1):(N)}}{\operatorname{argmin}}\{\alpha[l_{(1)}]\,V(\Omega[l_{(1)}]) + \sum\nolimits_{i=2}^{N}\alpha[l_{(i)}]V(\Omega[l_{(i)}|\Lambda_{(1):(i-1)}])\}. \quad (4)$$

In [7], $\alpha[l_j] = 1$ for all j. This reduces to searching the minimum overall search range. We find that unit computation cost is roughly proportional to the physical disk size needed to store the detector model; hence we set $\alpha[l_{(i)}]$ as the model disk size.

As in [7], we use a *greedy algorithm* for finding the ordering $\{l_{(1)}, ..., l_{(N)}\}$ that attempts to minimize the overall cost proceeds as follows:

Initialize $\Lambda = \phi$.
for *j=1,...,N* **do**
  $l_{(j)} = \arg\min_k \alpha[k]V(\Omega[k|\Lambda])$;
  Append $l_{(j)}$ to the ordered set Λ so that the new $\Lambda = l_{(1)}, ..., l_{(j)}$.
end

In other words, in each round one triggers the detector that yields the smallest computation.

It is easy to prove that the overall cost function in Eq.(4) can be reducible to a submodular function [7]. Optimizing submodular functions is in general NP-hard [14]. One must in principle evaluate $N!$ detector ordering patterns. Yet amazingly, *the greedy algorithm is guaranteed to find an ordered set Λ such that the invoked cost is at least 63% of its optimal value [7]!* It is worth emphasizing that the ordering found by the algorithm is data-dependent and determined on the fly.

4   Experimental Results

We present two sets of experimental results. The first is on NN matching for body region detection and the second about fast and accurate detection of one or multiple landmarks. The system runs on an Intel Xeon 2.33GHz CPU with 3GB RAM.

4.1   NN Matching for Body Region Detection

In this experiment, we use the NN matching for detecting body regions that need rough landmark locations by utilizing the holistic context. Our matching-based approach requires a database sufficiently large so that, given a query, the best match in the training database indeed covers the same body region(s) as the query. We collect 2,500 volumes annotated with 60 anatomical landmarks, including the left/right lung tops, aortic arch, femur heads, liver top, liver center, coccyx tip, etc. We use 500 volumes for constructing the training database and the remaining 2,000 volumes for testing. To ensure that

each query finds a good match, we construct our database of 100,000 volumes in a near-exhaustive manner: In each iteration, we randomly pick one of the 500 volumes and then randomly crop and slightly rotate it into a new volume before adding it to the database. The annotated anatomical landmark positions in the original volume are transformed accordingly.

	Median time (ms)	Average of Median errors (mm)
Baseline [7]	450	28.6
Our method	5	29.9
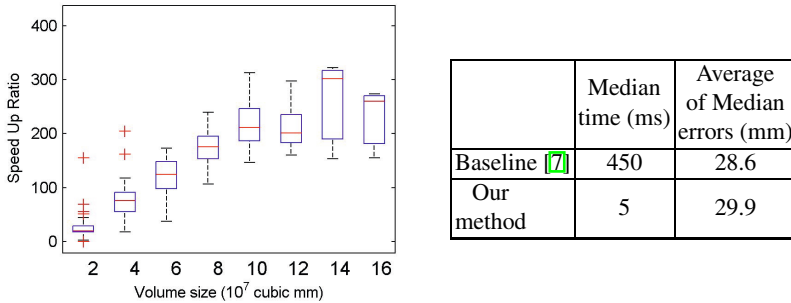
Fig. 2. The performance of detecting body regions using NN matching. The left plot shows the speed up ratio vs. the volume size.

Registration based methods are not applicable since the test volumes cover a large variety of body regions. If each region is detected separately say using [6], the total detection time is proportional to the number of regions, as detecting each region requires a scan over the whole volume. The work in [15] reports a detection time around 2000 ms for 9 landmarks, and median distance error around 22mm on a GPU (parallelized) implementation. The work in [7] has the highest accuracy and fastest speed, so we compare against this work in better detail. As in Fig. 2, our implementation of [7], which is tuned to a similar detection accuracy as shown in Table 1, has a detection time of 450ms for 6 landmarks that define the presence of right lung, skull, aorta, sternum, and liver; but the maximum time is 4.9sec, significantly larger than the median. This poses a problem for time critical subsequent tasks. The proposed method has a nearly constant detection time of 5 ms, achieving a speed-up of 90 times while maintaining similar detection accuracy. The speed-up is even more significant if more regions are of interest as our detection does not depend on the number of regions. Our NN matching code can be optimized and parallelized for faster speed. In general, a large detection error from NN matching, which is fine for body region detection purpose, is due to the large variability in the landmark appearance and its relative location to other landmarks.

Table 1. Median detection errors (mm) for 6 different landmarks that define 5 body regions

	Lung apex right	Skull base	Aortic root	Lung center	Sternum bottom	Liver center
Baseline [7]	24.1	31.9	23.2	20.6	37.3	35.2
Our method	27.1	19.1	35.8	24.5	35.1	37.9

4.2 Landmark Detection

When accurate positions are desired, we combine the NN matching with landmark detectors that exploit unitary context. Now each landmark detector only needs to search

in a local neighborhood around the rough position estimate given by the first stage, instead of searching in the whole volume. For detection of multiple landmarks, we further utilize pairwise spatial context for more improvements.

Detecting One Landmark. We consider detecting the liver top. In Fig. 3, the baseline approach uses a Probabilistic Boosting Tree (PBT) [6] to scan through the whole volume. Our method uses the PBT only to search in a local neighborhood. Evidently our method is much faster than the state-of-the-art due to the additional leverage of holistic context. A bigger volume yields more pronounced speedup (as large as 6-fold) as the use of holistic context breaks down the dependency on volume size.

	Median time (ms)	Average of Median errors (mm)
Baseline [6]	340	1.3
Our method	165	1.3
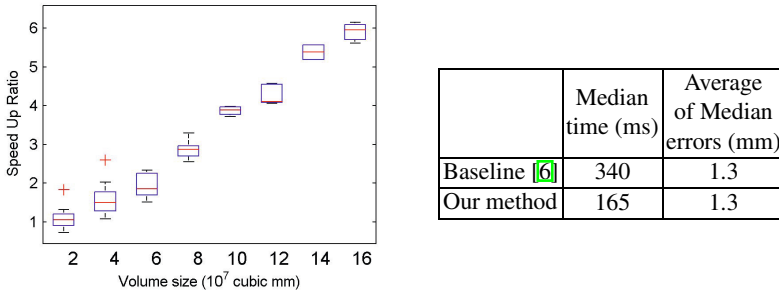
Fig. 3. The performance of accurately detecting the liver top

Detecting Multiple Landmarks. We further experiment accurately detecting 7 landmarks listed in Table 2 with three example landmarks of trachea bifurcation, liver bottom, and left kidney center shown in Fig. 4. Table 2 presents the mean detection error and the 95th percentile error that exhibits the robustness of the combined approach. The results in [7] are also included for comparison. We obtain better detection results except for the left kidney center, whose annotations are quite ambiguous, while consuming less time with a mean computation of 1.1s vs 1.3s for [7]. Due to space limitation, we omit the results of 16 other organs and anatomical structures.
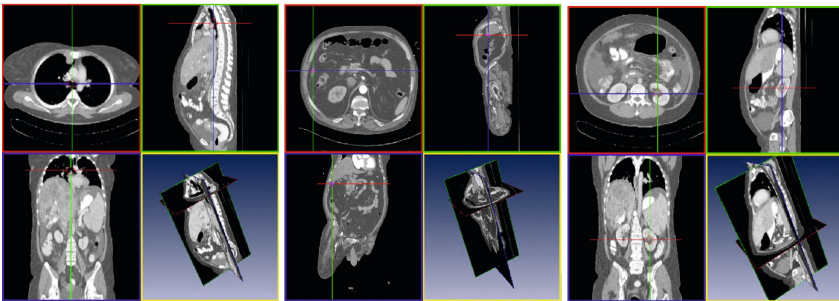
Fig. 4. Detected positions of trachea bifurcation, liver bottom, and left kidney center

Table 2. Errors (mm) in accurately detecting 7 different landmarks using NN matching and submodular optimization

(mm)	Mean	Q95	Mean [7]
Trachea bifurcation	**2.5**	4.5	2.8
Left Lung Top	**2.6**	6.0	3.2
Right Lung Top	**3.2**	8.5	3.7

(mm)	Mean	Q95	Mean [7]
Liver Top	**2.5**	4.0	2.9
Liver Bottom	6.4	30.5	n.a.
Left Kidney Center	8.4	50.7	**6.3**
Right Kidney Center	**6.4**	39.2	7.0

5   Conclusions

In this work we have introduced a fast and accurate method to detect landmarks in 3D CT data. Our method outperforms the state-of-the-art methods in detection speed with improved and comparable accuracy. The improvements arise from the leverage of holistic contextual information in the medical data via the use of an approximate NN matching to quickly identify the most similar database volume and transfer its landmark positions and the exploitation of unitary and pairwise context via a submodular formulation that aims to minimize the total computation for detecting landmark(s) and renders itself tocs a computationally efficiently greedy algorithm. Our method has been successively validated on a database of 2,500 CT volumes. In future we will extend it to different modalities such as MRI.

References

1. Rangayyan, R., Banik, S., Rangayyan, R., Boag, G.: Landmarking and Segmentation of 3D CT Images. Morgan & Claypool Publishers (2009)
2. Crum, W.R., Phil, D., Hartkens, T., Hill, D.: Non-rigid image registration: theory and practice. British Journal of Radiology 77, 140–153 (2004)
3. Pauly, O., Glocker, B., Criminisi, A., Mateus, D., Möller, A.M., Nekolla, S., Navab, N.: Fast Multiple Organ Detection and Localization in Whole-Body MR Dixon Sequences. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part III. LNCS, vol. 6893, pp. 239–247. Springer, Heidelberg (2011)
4. Seifert, S., Kelm, M., Moeller, M., Mukherjee, S., Cavallaro, A., Huber, M., Comaniciu, D.: Semantic annotation of medical images. In: SPIE Medical Imaging (2010)
5. Viola, P., Jones, M.: Robust real-time face detection. Intl. J. of Comp. Vis. 57, 137–154 (2004)
6. Tu, Z.: Probabilistic boosting-tree: Learning discriminative models for classification, recognition, and clustering. In: Proc. ICCV, pp. 1589–1596 (2005)
7. Liu, D., Zhou, S.K., Bernhardt, D., Comaniciu, D.: Search strategies for multiple landmark detection by submodular maximization. In: Proc. CVPR (2010)
8. Zhan, Y., Zhou, X.S., Peng, Z., Krishnan, A.: Active Scheduling of Organ Detection and Segmentation in Whole-Body Medical Images. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 313–321. Springer, Heidelberg (2008)
9. Seifert, S., et al.: Hierarchical parsing and semantic navigation of full body CT data. In: SPIE Medical Imaging (2009)
10. Zhou, S.K.: Shape regression machine and efficient segmentation of left ventricle endocardium from 2D B-mode echocardiogram. Med. Image Anal. 14, 563–581 (2010)

11. Izard, C., Jedynak, B., Stark, C.E.L.: Spline-Based Probabilistic Model for Anatomical Landmark Detection. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4190, pp. 849–856. Springer, Heidelberg (2006)
12. Torralba, A.: Contextual priming for object detection. Intl. J. Comp. Vis. 53, 169–191 (2003)
13. Jegou, H., Douze, M., Schmid, C.: Product quantization for nearest neighbor search. IEEE Trans. on Pattern Analysis and Machine Intelligence 33, 117–128 (2011)
14. Lovasz, L.: Submodular Functions and Convexity, pp. 235–257. Springer (1983)
15. Criminisi, A., Shotton, J., Bucciarelli, S.: Decision forests with long-range spatial context for organ localization in CT volumes. In: MICCAI Wksp. on Prob. Models for MIA (2009)

Integration of Local and Global Features
for Anatomical Object Detection in Ultrasound

Bahbibi Rahmatullah[1], Aris T. Papageorghiou[2], and J. Alison Noble[1]

[1] Institute of Biomedical Engineering, Dept. of Engineering Science, University of Oxford
[2] Nuffield Dept. of Obstetrics and Gynaecology, University of Oxford, United Kingdom
{bahbibi.rahmatullah,alison.noble}@eng.ox.ac.uk,
aris.papageorghious@obs-gyn.ox.ac.uk

Abstract. The use of classifier-based object detection has found to be a promising approach in medical anatomy detection. In ultrasound images, the detection task is very challenging due to speckle, shadows and low contrast characteristic features. Typical detection algorithms that use purely intensity-based image features with an exhaustive scan of the image (sliding window approach) tend not to perform very well and incur a very high computational cost. The proposed approach in this paper achieves a significant improvement in detection rates while avoiding exhaustive scanning, thereby gaining a large increase in speed. Our approach uses the combination of local features from an intensity image and global features derived from a local phase-based image known as feature symmetry. The proposed approach has been applied to 2384 two-dimensional (2D) fetal ultrasound abdominal images for the detection of the stomach and the umbilical vein. The results presented show that it outperforms prior related work that uses only local or only global features.

Keywords: Ultrasound, Local phase, Monogenic signal, Feature symmetry, Haar features, AdaBoost, Anatomical object detection.

1   Introduction

Ultrasound imaging is considered the simplest, least expensive and most widely used imaging modality in the field of obstetrics. Standard fetal biometric measurements from 2D ultrasound have been extensively used to estimate the gestational age of the fetus, to track fetal growth pattern, to estimate fetal weight and to detect abnormalities. Typically, fetal biometry is determined from standardized ultrasound planes taken from the fetal head, abdomen and thigh. Fetal growth is then assessed from these measurements by using population-based growth charts. The acquisition of standard image planes where these measurements are taken from is crucial to allow for accurate and reproducible biometric measurements, and also to minimize inter- and intra-observer variability.

Detection of medical anatomic structure plays an important role in medical image understanding and application. In our application, the detection of the important anatomical landmarks is one of the pre-defined criteria for qualitative scoring of a

fetal biometric image [1]. For example, the correct abdominal section for taking the abdominal circumference measurement (one of the important fetal biometric measures) should demonstrate the main anatomic landmarks: the stomach (SB) and the umbilical vein (UV). Results from the detection can also provide valuable initialization information for applications such as segmentation and registration. Therefore, the methodology presented in this paper is general and could be adapted for the detection of other anatomical features in medical imaging.

The use of classifier-based object detection with bounding boxes [2] which gained its popularity for generic object detection in natural images has found to be promising in medical anatomy detection [3]. The standard approach is to train the binary classifier by discriminating the anatomic structure of interest from the background. Then, exhaustive scan using the sliding window technique is performed (for all possible translations and a sparse set of scales) to find the anatomical object in the query image. In other words, a classifier is applied to all sub-windows within an image and takes the maximum of the classification score as an indication of the presence or absence of an object. One inherent disadvantage of this approach is the significant increase in computational cost, because of the large number of candidate sub-windows.

In [4], Rahmatullah et al. proposed the use of Haar features for the detection of important anatomical landmarks in fetal abdominal images. However, the local detector was applied on the entire image, resulting in an inefficient computational time of six seconds for an image size of 1024 by 768 pixels. In order to avoid the exhaustive search, our proposed approach includes global features that are formulated to predict the likely location of the object. It is important that the features used are detectable even under changes in image scale, noise and contrast. Moreover, in the case of fetal ultrasound images, the fetus shape and anatomy varies during pregnancy and the image quality diminishes with gestational age (GA) with typically stronger artifacts appearing in the image towards the end of pregnancy. The anatomical object can also appear at any pose in the images.

Encouraging results in other ultrasound-based applications (for example [5,6]) have shown that a local phase based approach outperforms the conventional intensity-based approach for feature detection in ultrasound images . In this paper, we propose the feature symmetry (FS) measure [7] derived from the local phase-based method as a global (coarse) feature for object detection. A local (fine) detector is then applied only to the locations deemed probable by the global features. To the best of our knowledge, the current study is the first to demonstrate the efficiency and the significance of integrating the feature symmetry in a machine learning framework for the detection of anatomical features in ultrasound images.

2   Local Phase Based Feature Measurement

Local phase information can be obtained by convolving a 1D signal with a pair of band-pass quadrature filters. A common choice of quadrature filters is the log-Gabor filter, which has a Gaussian transfer function when viewed on the logarithmic frequency scale. The log-Gabor filter has a transfer function of the form:

$$G(\omega) = exp\left(-\frac{log^2(\omega/k)}{2log^2(\sigma_\omega)}\right) \tag{1}$$

where k is the centre frequency of the filter and $0 < \sigma_\omega < 1$ is related to the spread of the frequency spectrum in a logarithmic function. Log-Gabor filters allow arbitrarily large bandwidth filters to be constructed while still maintaining a zero DC component in the even-symmetric filters.

The monogenic signal was introduced to calculate local phase in N-dimensional signals [8]. The monogenic signal is generated using the Riesz transform. The spatial representations of these filters are: $h_1(x,y) = -x/(2\pi(x^2 + y^2)^{\frac{3}{2}})$ and $h_2(x,y) = -y/(2\pi(x^2 + y^2)^{\frac{3}{2}})$. The image $I(x,y)$ is first convolved with an even isotropic band-pass filter $b(x,y)$ that produces the even (symmetric) component of the monogenic signal: $even(x,y) = I_b(x,y) = b(x,y) * I(x,y)$. The bandpassed image $I_b(x,y)$ is then convolved with the Riesz filter to produce the two odd (anti-symmetric) components: $odd_1(x,y) = h_1(x,y) * I_b(x,y)$ and $odd_2(x,y) = h_2(x,y) * I_b(x,y)$. The monogenic signal $I_M(x,y)$ of $I(x,y)$ is often expressed as $I_M(x,y) = [I_b(x,y), h_1(x,y) * I_b(x,y), h_2(x,y) * I_b(x,y)]$.

The stomach and the umbilical vein in fetal ultrasound images, typically appear as dark blobs, with non-uniform intensity and sizes in different scans. Therefore we propose the use of the multi-scale feature symmetry (FS) measure [7] for computing the phase congruency for these two objects, which is defined as:

$$FS(x,y) = \sum_s \frac{\lfloor[|even_s(x,y)| - |odd_s(x,y)|] - T_s\rfloor}{\sqrt{even_s(x,y)^2 + odd_s(x,y)^2} + \varepsilon} \tag{2}$$

where s represents the scale of the band-pass filter, ε is a small positive constant that avoids division by zero, $\lfloor.\rfloor$ operator denotes zeroing of any negative values, and T_s is scale specific noise compression term defined similarly in [5] as $T_s = exp\left(mean(log(\sqrt{even_s(x,y)^2 + odd_s(x,y)^2}))\right)$.

For band-pass filtering with log-Gabor filter, the following parameters produced the best empirical results: $\sigma_\omega = 0.50$ and 3 scales of filter wavelength: [250 150 50] pixels. The filter wavelength scales were selected from the coarse, medium, and fine ranges that would produce local phase images as shown in Fig. 1 (a)-(c). FS values vary from a maximum of 1, indicating a very significant feature, down to 0 indicating no significance. Based on the observation of the validation set, we set the significance value of 0.35. Connected component labeling was then applied to the FS image and the local discriminative detector was applied around the center location of the labeled component. The same FS image was used for both the stomach and the umbilical vein. In our validation set, the size of the stomach component in the labeled image usually varied from 800 to 2500 pixels and the umbilical vein from 300 to 1000 pixels. There might be overlap of potential components for the stomach and the umbilical vein but the task to distinguish them is passed to the local detector trained with the local features. Fig. 1 shows the prediction of the candidate locations of objects based on the global feature symmetry measure.
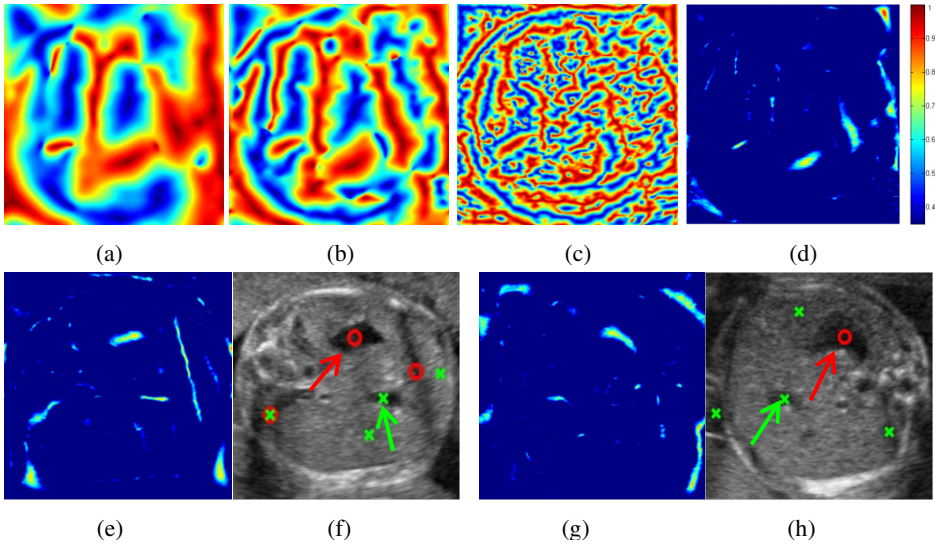
Fig. 1. The top row shows the local phase images produced using (a)-(c) coarse to fine filter scales and (d) the feature symmetry map generated using these scales. The bottom row shows the feature symmetry maps (e) and (g) with significant features (>0.35) that were used to produce the candidate locations for the stomach (*red circles*) and the umbilical vein (*green crosses*) shown superimposed on the original intensity image in (f) and (h). The *red* and *green arrows* denote the correct positions for the stomach and the umbilical vein, respectively.

3 Local Features and Training

The standard approach to local object detection is to classify each image sub-window as foreground (containing the object) or background. In this case, there are two main decisions to be made: what kind of local features to extract from each sub-window, and what kind of classifier to apply to this feature vector.

We propose to use features derived from Haar wavelets for representing the information in sub-window (local) image region. Haar features had been proven to effectively encode the domain knowledge for different types of visual patterns including objects in ultrasound images [3]. Extraction is achieved with high speed of computation due to the use of the integral image [2]. The features are trained as a local classifier using AdaBoost [9], an intuitive and proven effective method in object detection problem. It forms a strong hypothesis through linear combination of weak classifiers that are derived from a pool of extracted features. The training process involves modification of the weight distributions based on the previous classification error in order to focus on the more difficult training samples, thus driving down the classification error. Advantages of AdaBoost algorithm are that it has no parameters to tune other than the number of iterations and the most representative features are automatically selected during the training process, requiring no additional experiments.

Our training data for the classifier is created as follows. For positive training samples, we cropped image regions that contain the anatomical object. The negative samples were extracted randomly from the background and also from images that does not contain SB or UV. The cropped sub-windows were normalized to 100x100. From the training dataset, we extracted the following separate feature sets:

    a. *Local features*: Haar features extracted from intensity image
    b. *Global features*: Unary features acquired from the feature symmetry (FS) image.

300 rounds of boosting were performed on each feature set. This is based on the performance on the validation set. The resulting first five features chosen for each object are shown in Fig. 2.

(a)   (b)
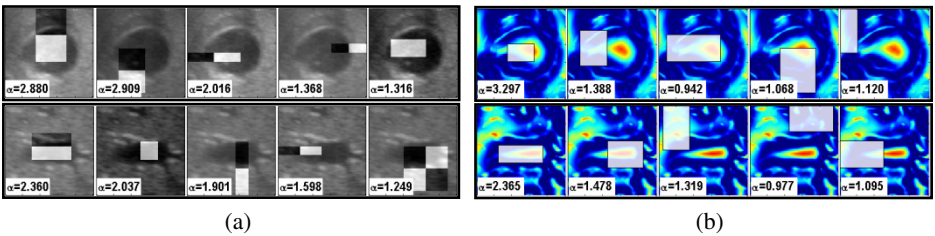
Fig. 2. The first five selected features by AdaBoost for the stomach (*top row*) and the umbilical vein (*bottom row*) detection are shown superimposed on example images from the training set. The features are calculated by summing the value in white regions and subtracted with the totals in grey regions (if any). (a) Local features from intensity image and (b) global features from feature symmetry (FS) image are shown with each feature classifier weight (α).

4   Experimental Setup and Results

4.1   Data Acquisition

Fetal abdominal images for this work were randomly selected from a clinical study database where data has been obtained from over 4000 healthy pregnant women at low risk of impaired fetal growth who are scanned up to six times from <14_{+0} weeks to term. All women were screened at study entry with particular focus on excluding known risk factors for IUGR (e.g. smoking, chronic illness) and over-growth (e.g diabetes). Multiple pregnancy or major fetal abnormality were excluded from the study. All ultrasound examinations were performed using a Philips HD9 ultrasound machine with a 2-5MHz 2D probe by ultrasonographers trained to follow standardized data acquisition procedures. All images were saved in a DICOM format with a size of 1024 x 768 pixels. Details of datasets is shown in Table 1 where images were labeled and divided (no overlaps) after consultation with trained sonographers.

Table 1. Details of the number of positive (+) and negative (-) images in the training, validation and testing datasets

	Train+	Train-	Valid+	Valid-	Test+	Test-
SB	633	2073	50	50	2283	101
UV	224	851	50	50	2284	100

4.2   Results and Performance Analysis

The experiment was implemented in MATLAB running on a Pentium Xeon® 3.4 GHz machine with 3GB of memory. We experimented with three different detection methods: "Local", "Global" and "Hybrid". In the "Local" and the "Global" methods, we exhaustively scanned the image at multiple scales using the sliding window method. Features from each sub-window were extracted and classified according to its trained model (local Haar features or global FS). For "Hybrid", the detector trained with the local features was applied at the probable locations and scales predicted by the global features (FS). We empirically set that a ground-truth object is considered detected if 75% of its area is covered by the output boxes.

We compared the performance of the three methods using ROC curves, as shown in Fig. 3 and summarize the performance metric; area under the curve (AUC), accuracy, sensitivity, specificity and execution time, shown in Table 2. From the results, we found that the method using global and local features is computationally efficient and eliminates many false positives caused by using local features alone, and these are illustrated qualitatively in Fig. 4. The umbilical vein detection is a harder problem hence the lower AUC value and accuracy compared to the stomach detection. This is due to the presence of other similar looking blood vessels in the abdomen causing false positive in the detection. However the detection accuracy is increased by 9.75% with the integration of global feature with the local features in the "Hybrid" method.

Table 2. The performance of three different methods in the detection of the stomach and the umbilical vein in fetal abdominal images

	AUC		Accuracy (%)		Sensitivity (%)		Specificity (%)		Mean Execution
	SB	UV	SB	UV	SB	UV	SB	UV	Time (secs)
Local	0.80	0.57	78.94	62.80	60.84	54.59	96.04	71.00	10.27
Global	0.71	0.53	69.28	57.99	59.35	33.98	79.21	82.00	10.65
Hybrid	0.88	0.75	82.75	72.55	66.49	57.09	99.01	88.00	0.94
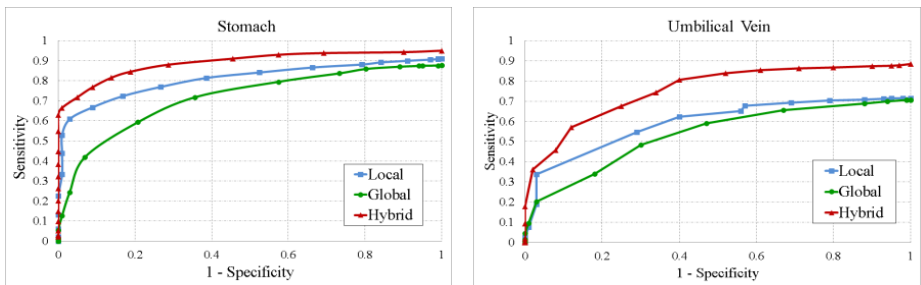
Fig. 3. ROC plot for the detection of the stomach (*left*) and the umbilical vein (*right*). We see that combining the global and local features improves detection performance.

    (a) FS map   (b) Stomach detection   (c) FS map   (d) UV detection
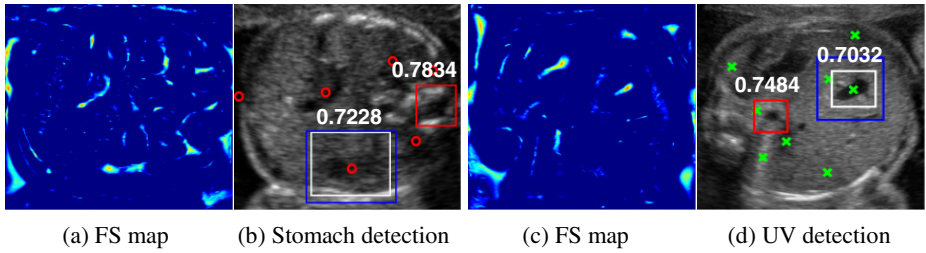
I) False positive results by "*Local*" method (*red boxes*) corrected with true positives by "*Hybrid*" method (*blue boxes*) along with the detection scores.

    (a) FS map   (b) Stomach detection   (c) FS map   (d) UV detection
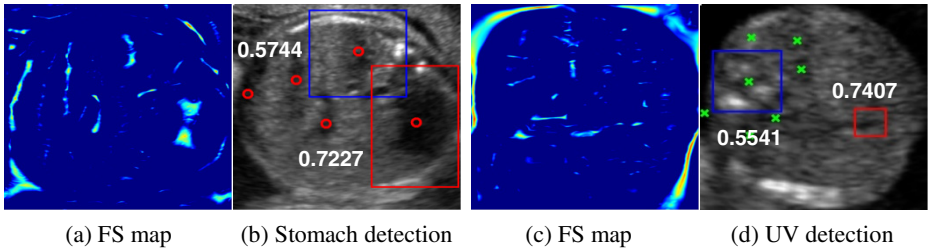
II) False positive results (high detection scores) by "*Local*" method (*red boxes*) corrected with true negatives (low detection scores) by "*Hybrid*" method (*blue boxes*).

    (a) FS map   (b) Stomach detection   (c) FS map   (d) UV detection
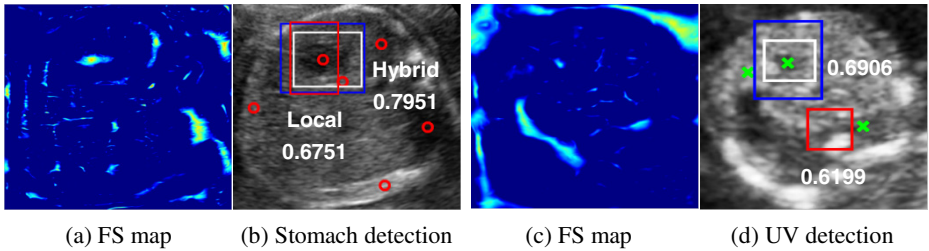
III) False negative results (low detection scores) by "*Local*" method (*red boxes*) corrected with true positives (high detection scores) by "*Hybrid*" method (*blue boxes*).

    (a) FS map   (b) Stomach detection   (c) FS map   (d) UV detection
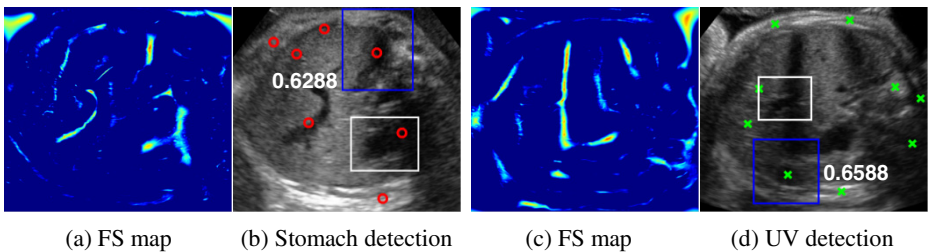
IV) Misdetections which occurred mostly in late GA images. The correct location of the objects could not be identified through the FS map due to major shadowing effect over the objects.

Fig. 4. Examples of the results achieved using "Hybrid" method. White box represents the ground truth. (a) and (c) show the FS maps that input the candidate locations for stomach (*red circles*) and UV (*green crosses*), respectively.

5   Conclusion

This paper presents a novel feature set for the detection of anatomical objects in fetal ultrasound image. We integrated the unary features extracted from local phase image global feature within a machine learning framework that trains a local classifier using local Haar features. This provides a computationally cheap step before invoking a local object detector to be applied in plausible locations and scales. The proposed method exhibits good generalization capability when tested on 2384 images with an accuracy of 82.75% and 72.55% for the detection of the stomach and the umbilical vein, respectively. In average, it runs 9 times faster than the typical local object detector with the sliding window approach. Our future works will focus on the extension of the proposed method for the detection of objects in other fetal biometry scan areas and in 3D ultrasound volumes.

References

1. Salomon, L.J., et al.: Feasibility and reproducibility of an image-scoring method for quality control of fetal biometry in the second trimester. Ultrasound in O&G 27, 34–40 (2006)
2. Viola, P., Jones, M.J.: Robust Real-Time Face Detection. International Journal of Computer Vision 57, 137–154 (2004)
3. Karavides, T., Leung, K.Y.E., Paclik, P., Hendriks, E.A., Bosch, J.G.: Database guided detection of anatomical landmark points in 3D images of the heart. In: 2010 IEEE International Symposium on Biomedical Imaging, Rotterdam, pp. 1089–1092 (2010)
4. Rahmatullah, B., Sarris, I., Papageorghiou, A., Noble, J.A.: Quality control of fetal ultrasound images: Detection of abdomen anatomical landmarks using AdaBoost. In: 2011 IEEE International Symposium on Biomedical Imaging, Chicago, IL, pp. 6–9 (2011)
5. Rajpoot, K., Vicente, V.V., Noble, J.A.: Local-phase based 3D boundary detection using monogenic signal and its application to real-time 3-D echocardiography images. In: 2009 IEEE International Symposium on Biomedical Imaging, Boston, MA, pp. 783–786 (2009)
6. Grau, V., Noble, J.A.: Adaptive Multiscale Ultrasound Compounding Using Phase Information. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3749, pp. 589–596. Springer, Heidelberg (2005)
7. Kovesi, P.: Symmetry and Asymmetry From Local Phase. In: Tenth Australian Join Conference on Artificial Intelligence, Brisbane, pp. 185–190 (1997)
8. Felsberg, M., Sommer, G.: The monogenic signal. IEEE Transactions on Signal Processing 49, 3136–3144 (2001)
9. Freund, Y., Schapire, R.E.: A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. Journal of Computer & System Sciences 55, 119–139 (1997)

Spectral Label Fusion

Christian Wachinger[1,2] and Polina Golland[1]

[1] Computer Science and Artificial Intelligence Lab, MIT
[2] Martinos Center for Biomedical Imaging, Harvard Medical School

Abstract. We present a new segmentation approach that combines the strengths of label fusion and spectral clustering. The result is an atlas-based segmentation method guided by contour and texture cues in the test image. This offers advantages for datasets with high variability, making the segmentation less prone to registration errors. We achieve the integration by letting the weights of the graph Laplacian depend on image data, as well as atlas-based label priors. The extracted contours are converted to regions, arranged in a hierarchy depending on the strength of the separating boundary. Finally, we construct the segmentation by a region-wise, instead of voxel-wise, voting, increasing the robustness. Our experiments on cardiac MRI show a clear improvement over majority voting and intensity-weighted label fusion.

1 Introduction

Label fusion has gained much popularity in medical image segmentation. It benefits from prior information in a form of images previously labeled by experts. Instead of summarizing the data as a probabilistic atlas [2], label fusion algorithms maintain and use all the labeled images [6,9,11]. In various comparisons, label fusion outperforms alternative atlas-based segmentation strategies when the anatomical variability is too large to be represented by mean statistics [6,9].

In computer vision, spectral techniques, such as normalized cuts [12], are commonly employed for segmentation. Central to these methods is the quantification of pairwise similarities between points in the image, which serve as weights for the graph Laplacian. Earlier work computed the similarity by comparing intensity values [12]. Further studies introduced intervening contours, which relate the similarity between two locations to the existence of a boundary separating them [5]. In order to extract the boundary, image edges and textons are calculated to combine contour and texture cues [8]. The spectral framework constructs a globally optimal partitioning based on these local measures. In a recent comparison [1], a multiscale version of normalized cuts outperformed other contour detectors and produced excellent segmentation results on natural images.

The mentioned spectral algorithms, as well as active contour and level set techniques, obtain impressive results by solely considering image cues. In contrast, label fusion segmentation is mainly based on intensity differences between the images, but the contour information in the test image is rarely exploited. Our method addresses exactly this issue and closes the gap between label fusion

(a) Label fusion   (b) Image contour based   (c) Spectral label fusion
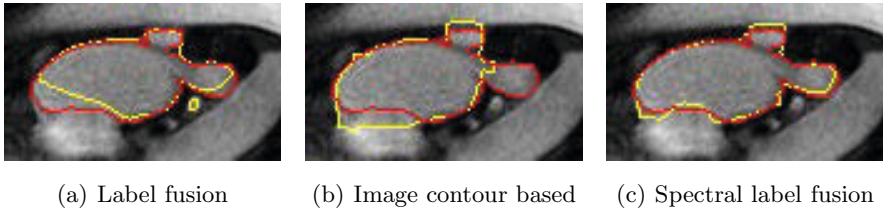
Fig. 1. Example segmentations of the left atrium of the heart. Automatic segmentation is shown in yellow, expert labeling is shown in red.

techniques and spectral segmentation approaches. We combine image contour cues and label information in the graph Laplacian to produce contours that jointly depend on the atlas and the test image. Fig. 1 illustrates that this is beneficial. Intensity-weighted label fusion [11] leads to an undersegmentation of the left atrium of the heart (Fig. 1(a)). We observed this behavior in many cases. Using image contours only faces challenges in identifying the correct boundaries, especially for smooth intensity transitions (Fig. 1(b)). As a result, certain parts are oversegmented while others, e.g., the veins, are cut off. The segmentation with spectral label fusion, which fuses image and label information, leads to the most accurate segmentation. The proposed algorithm constitutes a novel approach to integrate prior information in spectral segmentation and leverages state-of-the-art contour and region extraction methods to enhance label fusion.

1.1 Clinical Motivation

The proposed approach is well suited for delineating structures of high variability, which are challenging to segment with atlas-based techniques due to difficulties of inter-subject registration. With the integration of local image cues, the method is less prone to alignment errors. We focus on the left atrium of the heart, which exhibits large variations in the shape of the cavity and in the number and location of pulmonary veins [4]. Its segmentation is of significant clinical relevance, because one of the most common heart conditions is atrial fibrillation [4]. In atrial fibrillation, the left atrium no longer pumps blood into the left ventricle efficiently but instead quivers in an abnormal way. The common treatment is radio-frequency catheter ablation of ectopic foci [7]. Accurate segmentation of the left atrium and its pulmonary venous drainages on contrast-enhanced magnetic resonance angiography (MRA) images is essential for planning and evaluating ablation procedures.

1.2 Related Work

In addition to prior work reviewed above, our work is also related to a relatively new approach for brain image segmentation based on non-local means filtering [3,10]. The method compares image patches between the test subject

and the atlas images, with voting that depends on intensity differences between these patches. In our application this technique seems less promising because we work with contrast-enhanced MRA images, characterized by substantial intensity variations. In [14], a local search on patches is performed to improve the results after the atlas-based segmentation. Instead of refining the segmentation through the integration of image information in a post-processing step, we consider image and label information jointly in the graph Laplacian. Moreover, we work on image contours that have advantages over comparing intensities [5].

Recently, an algorithm based on graph cuts has been demonstrated for the refinement of atlas propagation [13,15]. In contrast to this method, we apply techniques from spectral clustering and combine gradient and texture cues, leading to a region-based voting on an oversegmentation of the image. In [7], the segmentation of the left atrium is obtained by extracting the blood pool with intensity thresholding, which is sensitive to intensity variations. Intensity-weighted label fusion achieved accurate results for the segmentation of the left atrium in [4]. We therefore treat this label fusion technique as a baseline for comparison.

2   Method

Spectral label fusion consists of three steps, as illustrated in Fig. 2. The first step extracts the boundaries from the image and label map, joins them in the spectral framework, and produces weighted contours. In the second step, these contours give rise to regions, partitioning the image. In the third step, we assign a label to each region based on the input label map, producing the final segmentation. We formulate segmentation as a binary labeling problem; for a multi-label problem the same procedure is repeated for each label.

2.1   Input Data

The new image I to be segmented and the probabilistic label map $\hat{L}$ are the inputs to the algorithm. Any atlas-based approach, parametric or non-parametric, can be employed to create the label map, which serves as atlas-based label prior. In this work, we adapt the label fusion approach and register each of the images in the training set to image I with a variant of the diffeomorphic demons algorithm, iteratively fitting a polynomial transfer function to compensate for intensity differences in the MRA images [4]. We use $\mathcal{I} = \{I_1, \ldots, I_n\}$ to denote the transformed and intensity corrected training images. We apply the estimated warps to the labels, leading to propagated labels $\mathcal{L} = \{L_1, \ldots, L_n\}$. Following the formulation in [4,11], we compute the MAP labeling at location x:

$$\hat{L}(x) = \arg\max_l p(L(x) = l, I(x)|\mathcal{L}, \mathcal{I}) = \arg\max_l \sum_{i=1}^{n} p(L(x) = l|L_i) \cdot p(I(x)|I_i).$$
(1)

The label likelihood $p(L(x) = l|L_i) \propto \exp(\kappa D_i^l)$ serves as soft vote, weighted by the image likelihood $p(I(x)|I_i) \propto \exp(-\frac{(I(x)-I_i(x))^2}{2\sigma^2})$. D_i^l is the signed distance transform of label l in the warped label map L_i. We set $\kappa = 1.5$ and $\sigma^2 = 1$.
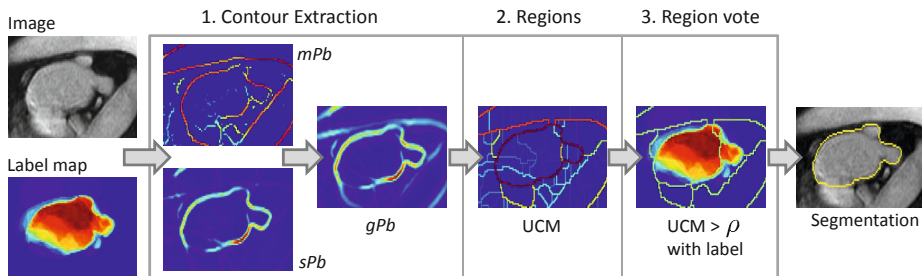
Fig. 2. Overview of spectral label fusion

2.2 Contours

We employ the concept of intervening contours to calculate the weights in the graph Laplacian, which leads to better results than the comparison of intensity values [5]. The first step is to estimate the probability $Pb(x)$ of a contour in each slice of the image I at location x. We choose to employ the oriented gradient signal [1]. The method robustly estimates the image gradient by calculating the χ^2 distance between the histograms of two half-discs at each location for various orientations. Depending on the size of the disc, we obtain gradient estimates on multiple scales. To quantify the texture in the image, we calculate textons by convolving the image with 17 Gaussian derivative and center-surround filters [8]. We obtain 64 different textons with a K-means clustering in the 17-dimensional space. The image and texton gradients of multiple scales are added, resulting in the multiscale contour probability $mPb(x)$ [1], as illustrated in Fig. 2. We use mPb to calculate weights between points i and j in the image, following the concept of intervening contours by identifying the maximum along the line $\overline{ij}$:

$$W_{ij}^I = \exp\left(-\max_{x \in \overline{ij}}\{mPb(x)\}\right). \tag{2}$$

Analogously, we estimate the probability of a contour in the label map $\hat{L}$, denoted by lPb, and derive the weights

$$W_{ij}^L = \exp\left(-\max_{x \in \overline{ij}}\{lPb(x)\}\right). \tag{3}$$

For computational efficiency the weights are only calculated within a radius $r = 5$, yielding sparse matrices [1].

Once the image and label weights are computed, we combine them in the joint weight matrix

$$W = W^I + W^L. \tag{4}$$

The generalized eigenvalue decomposition $(D - W)\mathbf{v} = \lambda D\mathbf{v}$ gives rise to eigenvectors $\mathbf{v}_k$ that partition the image based on image and label cues, with

$D_{ii} = \sum_i W_{ij}$. We consider the 16 eigenvectors corresponding to the smallest non-zero eigenvalues λ_k. Instead of performing a K-means clustering in the 16 dimensional space [12], which tends to break up uniform regions [1], we obtain the spectral boundary sPb by summing up the gradients of the eigenvectors $sPb = \sum_{k=1}^{16} \frac{1}{\sqrt{\lambda_k}} \cdot \nabla \mathbf{v}_k$. While mPb contains responses to all edges, sPb only shows the most salient structures in the image. The global contour is obtained by combining both maps to take advantage of the two characteristics, i.e., $gPb = mPb + \gamma \cdot sPb$, where γ is a weighting constant.

2.3   From Contours to Regions

In the second step, we use the extracted contours gPb to partition the image into regions. These regions form an oversegmentation of the image. Consequently, the left atrium and veins do not correspond to one but several regions. The size of these regions is subject to a trade-off. Large regions provide stability in the face of registration errors, but they are also more likely to miss the actual organ boundary. To enable adaptive region size selection, we use the strength of the extracted boundaries to build a hierarchical segmentation. At the lowest level, we have the finest partition of the image, and the higher levels contain larger regions implied by stronger contours.

Specifically, we use the oriented watershed transform to create the finest partition [1]. We experimented with the watershed in 2D and 3D. For the 3D watershed, we apply a 3D closing operation of the contours to prevent leakage in out-of-plane direction. Next, we employ the ultrametric contour map (UCM) [1] to represent the hierarchical segmentation, illustrated in Fig. 2. We clearly see that the strong boundary of the atrium appears in the UCM. We select the segmentation scale in the hierarchy by thresholding with parameter ρ.

2.4   Voting on Regions

One of the limitations of the current label fusion framework [11] is the assumption of independence of voxel samples, which is generally not justified for medical images. It is more appropriate to consider independence with respect to a local neighborhood, applying the Markov property in the derivation. Crucial is the selection of image-specific neighborhoods that capture the relevant information. The regions constructed in the previous step can naturally serve as such neighborhoods. We obtain the region-based MAP estimation by considering the location x as random variable and marginalizing over it

$$\tilde{L}^R = \arg\max_l \sum_{x \in R} p(L^R = l | \mathcal{L}, x) p(I^R | \mathcal{I}, x) p(x) \tag{5}$$

$$= \arg\max_l \sum_{x \in R} \sum_{i=1}^{n} \exp\left(\kappa D_i^l(x) - \frac{(I(x) - I_i(x))^2}{2\sigma^2} \right), \tag{6}$$

where L^R and I^R denote the region R in the label map and the intensity image, respectively. Since we aggregate the voting information over the entire region,
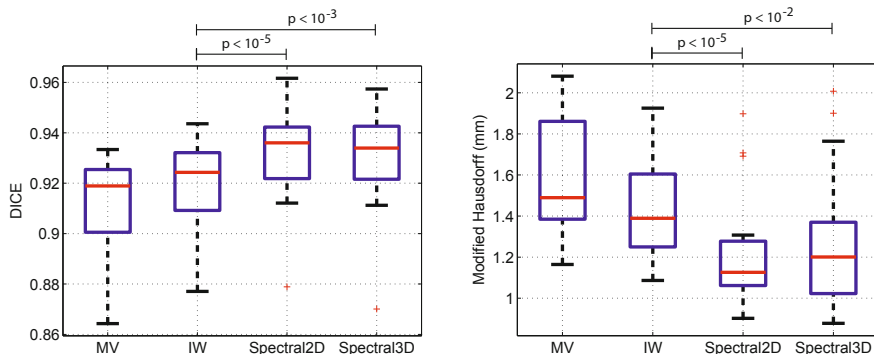
Fig. 3. Dice volume overlap (left) and modified Hausdorff distance (right). Red line indicates median, the boxes extend to the 25th and 75th percentiles, and the whiskers reach to the most extreme values not considered outliers (red crosses). Significance was evaluated using a single-sided paired t-test with IW as baseline.

the resulting method is more robust to registration errors and noise than a voxel-based approach.

3   Experiments

To evaluate the proposed method, we automatically segment the left atrium of the heart in a set of 16 electro-cardiogram gated (0.2 mmol/kg) Gadolinium-DTPA contrast-enhanced cardiac MRA images (CIDA sequence, TR=4.3ms, TE=2.0ms, $\theta = 40°$, in-plane resolution varying from 0.51mm to 0.68mm, slice thickness varying from 1.2mm to 1.7mm, $512 \times 512 \times 96$, -80 kHz bandwidth, atrial diastolic ECG timing to counteract considerable volume changes of the left atrium). The left atrium was manually segmented in each image by an expert. For all the experiments we set $\gamma = 2.5$, giving higher weight to the spectral component. We set $\rho = 0.2$ for the 2D and $\rho = 0$ for the 3D watershed after inspecting the UCM. We perform leave-one-out experiments by treating one subject as the test image and the remaining 15 subjects as the training set. We use the Dice score and the modified (average) Hausdorff distance between the automatic and expert segmentations as quantitative measures of segmentation quality. We compare our method to majority voting (MV) and intensity-weighted label fusion (IW) [11].

We apply a median filter in a $5 \times 5 \times 5$ window to the spectral segmentations. In-plane filtering has little effect; filtering improves the consistency and closes holes in out-of-plane direction. The application of the filter on MV and IW segmentations led to a deterioration on all subjects. We therefore present unfiltered results for these cases. Fig. 1 illustrates the results of IW and the 2D spectral label fusion, together with an approach that considers only image contours, $W = W^I$. Fig. 3 presents dice volume overlap and modified Hausdorff distance

(a) Majority Voting   (b) Intensity-weighted LF   (c) Spectral 2D
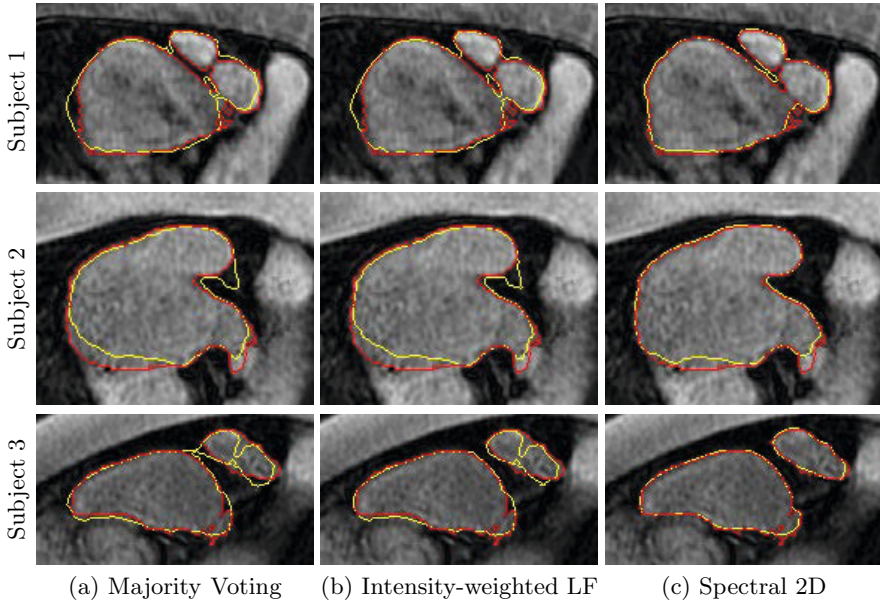
Fig. 4. Examples of automatic segmentation results for different subjects are shown in yellow. Manual delineations are shown in red.

for each algorithm. The improvements in segmentation accuracy between the proposed method and IW are statistically significant ($p < 10^{-5}$). At first glance, it may seem surprising that the 2D algorithm leads to better results than the 3D version. The anisotropic resolution of the data presents a challenge for extracting meaningful 3D regions. A further inspection of the results in each subject reveals that the values for IW are always better than the ones for MV, and the values of the 2D spectral method are consistently better than the ones for IW.

Fig. 4 illustrates the segmentation results for MV, IW, and 2D spectral fusion for several subjects. We see that spectral fusion better captures the organ boundary. This is supported by the clearly lower Hausdorff distances in Fig. 3. On the images for Subject 1 in Fig. 4, we observe that spectral label fusion achieves a better separation between the veins and atrium. This case is particularly challenging because the gap is small and registration errors of misaligning either the vein or the atrium lead to a closure. By integrating the image cues and voting on regions, we achieve a more accurate segmentation.

4   Conclusion

We presented spectral label fusion, a new approach for multi-atlas image segmentation. It combines the strengths of label fusion with advanced spectral segmentation. The integration of label cues into the spectral framework results in improved segmentation performance for the left atrium of the heart. The extracted image regions form a nested collection of segmentations and support a

region-based voting scheme. The resulting method is more robust to registration errors than a voxel-wise approach.

References

1. Arbelaez, P., Maire, M., Fowlkes, C., Malik, J.: Contour detection and hierarchical image segmentation. IEEE Trans. on Pat. Anal. Mach. Intel. 33(5), 898–916 (2011)
2. Ashburner, J., Friston, K.: Unified segmentation. NeuroImage 26(3), 839–851 (2005)
3. 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)
4. 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)
5. Fowlkes, C., Martin, D., Malik, J.: Learning affinity functions for image segmentation: Combining patch-based and gradient-based approaches. In: CVPR (2003)
6. Heckemann, R., Hajnal, J., Aljabar, P., Rueckert, D., Hammers, A.: Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. NeuroImage 33(1), 115–126 (2006)
7. Karim, R., Mohiaddin, R., Rueckert, D.: Left atrium pulmonary veins: segmentation and quantification for planning atrial fibrillation ablation. In: SPIE (2009)
8. Malik, J., Belongie, S., Leung, T., Shi, J.: Contour and texture analysis for image segmentation. Int. Journal of Computer Vision 43(1), 7–27 (2001)
9. Rohlfing, T., Brandt, R., Menzel, R., Russakoff, D., Maurer, C.: Quo vadis, atlas-based segmentation? In: Handbook of Biomedical Image Analysis, pp. 435–486 (2005)
10. Rousseau, F., Habas, P.A., Studholme, C.: A supervised patch-based approach for human brain labeling. IEEE Trans. Medical Imaging 30(10), 1852–1862 (2011)
11. Sabuncu, M., Yeo, B., Van Leemput, K., Fischl, B., Golland, P.: A Generative Model for Image Segmentation Based on Label Fusion. IEEE Transactions in Medical Imaging 29(10), 1714–1729 (2010)
12. Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(8), 888–905 (2000)
13. Song, Z., Tustison, N.J., Avants, B., Gee, J.C.: Integrated Graph Cuts for Brain MRI Segmentation. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4191, pp. 831–838. Springer, Heidelberg (2006)
14. Wang, H., Suh, J.W., Pluta, J., Altinay, M., Yushkevich, P.: Optimal Weights for Multi-atlas Label Fusion. In: Székely, G., Hahn, H.K. (eds.) IPMI 2011. LNCS, vol. 6801, pp. 73–84. Springer, Heidelberg (2011)
15. Wolz, R., Aljabar, P., Hajnal, J., Hammers, A., Rueckert, D.: LEAP: Learning embeddings for atlas propagation. NeuroImage 49(2), 1316–1325 (2010)

Multi-Organ Segmentation with Missing Organs in Abdominal CT Images

Miyuki Suzuki[1], Marius George Linguraru[2], and Kazunori Okada[1]

[1] Department of Computer Science, San Francisco State University
[2] Sheikh Zayed Institute for Pediatric Surgical Innovation
Children's National Medical Center, Washington DC
{miyukis,kazokada}@sfsu.edu, mlingura@cnmc.org

Abstract. Currently, multi-organ segmentation (MOS) in abdominal CT can fail to handle clinical patient population with missing organs due to surgical resection. In order to enable the state-of-the-art MOS for these clinically important cases, we propose 1) automatic missing organ detection (MOD) by testing abnormality of post-surgical organ motion and organ-specific intensity homogeneity, and 2) atlas-based MOS of 10 abdominal organs that handles missing organs automatically. The proposed methods are validated with 44 abdominal CT scans including 9 diseased cases with surgical organ resections, resulting in 93.3% accuracy for MOD and improved overall segmentation accuracy by the proposed MOS method when tested on difficult diseased cases.

1 Introduction

Multi-organ segmentation (MOS) has recently become popular toward improving overall segmentation accuracy when segmenting a set of organs located nearby, enabling comprehensive computer-aided diagnosis (CAD) of various multi-focal abdominal diseases [1–10]. In this paper, we investigate how such MOS can be extended to a patient population with missing organs due to surgical resections. Without considering this population, MOS cannot be applied to a number of important clinical applications such as follow-up studies of surgical treatment and cancer recurrence in abdomen. Despite this clinical importance, however, current MOS solutions are not designed to handle such cases with irregular anatomy. A common process in various MOS methods is to fit an atlas of normal organ anatomy to an image to be analyzed. When analyzing a case with missing organs, regardless of atlas formats (i.e., static [3], probabilistic [2, 4, 5, 8, 9], or geometric [4, 6–8, 10]), MOS can fail to segment other intact organs because of 1) mis-match of the atlas' part corresponding to the missing organs to nearby non-targets and 2) post-surgical organ shifts. Fig.1(a) illustrates such a failure case with a missing right kidney where the liver (red) shifted downward into the cavity caused by the removed kidney and a part of the liver was incorrectly identified as kidney (cyan).

Addressing the above issue, this paper presents two novel contributions to improve the current atlas-guided MOS solutions. First, we propose an automatic

(a)   (b)   (c)
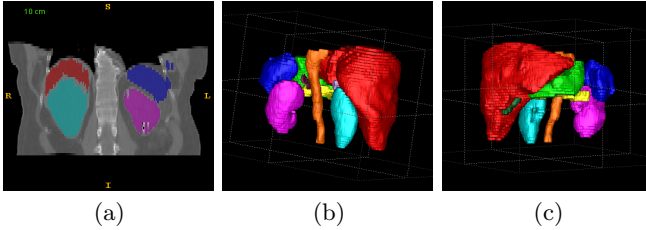
Fig. 1. Illustrative examples of a) segmentation failures (part of the liver is incorrectly labeled as kidney) and b,c) ten modeled organs. Red: liver, blue: spleen, cyan: r-kidney, magenta: l-kidney, yellow: pancreas, orange: aorta, dark green: gall bladder, purple: l-adrenal, lavender: r-adrenal, green: stomach.

missing organ detection (MOD) solution based on testing abnormality of data-driven features computed from the 4D spatio-intensity Gaussian mixture model (GMM) fitted to data. Three probabilistic features, capturing post-surgical organ motions, organ-specific intensity homogeneity, and their linear combinations, are proposed and compared. Such automatic MOD allows us to handle clinical scan data more robustly even when previous medical history information is missing or corrupted in patient record or DICOM tag [11]. Second, we present an atlas-guided MOS solution for 10 abdominal organs that automatically handles missing organs by incorporating the MOD solution to an atlas-guided maximum-a-posteriori (MAP) algorithm proposed in [4]. These proposed methods are validated with 44 abdominal CT scans, including 9 diseased cases with two common surgical resection procedures of splenectomy (spleen removal) and nephrectomy (kidney removal). Our experimental results demonstrate advantages of the proposed MOS method such that a correct MOD improves overall segmentation accuracy on average when dealing with the difficult diseased cases. The issue of handling missing organs in abdominal MOS is scarcely addressed in the literature. To the best of our knowledge, there is no previous studies that proposed an abdominal multi-organ segmentation with automatic missing organ handling.

2   Method

2.1   Atlas-Guided MAP Multi-Organ Segmentation

An atlas-guided MOS method proposed by Shimizu et al. [4] is adopted in this study as our base MOS method. This method employs the MAP estimation of organ label $l \in \{1, .., L\}$ over 4D spatio-intensity feature vector $\mathbf{v} = (x, y, z, I(x, y, z))$: $\hat{l} = \mathrm{argmax}_l p(\mathbf{v}|l)p(l)$. The prior $p(l)$ is modeled by a standard probabilistic atlas [2, 9]. The atlas $A_l(\mathbf{x}) \in [0, 1]$, $\mathbf{x} = (x, y, z)$, is built by registering K training images of normal anatomy to a fixed reference image I_R with a size-preserving affine registration then computing a probability map for each of L modeled organs by counting manually segmented organs. The likelihood $p(\mathbf{v}|l)$ is modeled by an extended GMM $p(\mathbf{v}) = \sum_{l=1}^{L} \sum_{n=1}^{N} \alpha_l(n) \mathcal{N}(\mathbf{v}; \mathbf{u}_l, \mathbf{\Sigma}_l)$ where N

denotes the number of voxels and the mixing weights $\alpha_l(n)$ are defined over each voxel n. To segment organs in a new image, the image is first registered to I_R using affine transformation followed by B-spline non-rigid registration [12]. From the K training images, a normal spatio-intensity model $(\mathbf{u}_{\mathbf{V}l}, \mathbf{\Sigma}_{\mathbf{V}l})$ for each organ l is also computed where $\mathbf{u}_{\mathbf{V}l}$ and $\mathbf{\Sigma}_{\mathbf{V}l}$ are the mean and covariance of feature vectors of the organ l. Initialized by this normal spatio-intensity model, $p(\mathbf{v})$ is fit to the new image using the EM-algorithm [13], yielding the patient-specific likelihood estimate $\{\hat{p}(\mathbf{v}|l)\}$. Additionally, the fitted GMM yields data-driven estimate of organ center and associated covariance $(\hat{\mathbf{u}}_{\mathbf{X}l}, \hat{\mathbf{\Sigma}}_{\mathbf{X}l})$ for each organ l.

2.2   Automatic Missing Organ Detection (MOD)

When fitting the GMM $p(\mathbf{v})$ to an image I^{mo} with missing organs, normal components in $p(\mathbf{v})$ corresponding to missing organs will be fitted to arbitrary non-target structures located nearby. Exploiting this observation, we propose a data-driven MOD by analyzing this EM model fitting error. Three probabilistic measures of missing organs, F_l, G_l, and H_l, are derived by testing abnormality of organ features estimated from the GMM fitting result with respect to respective normal models, as described below.

The first measure F_l indicates the probability of organ l to be missing by quantifying how abnormal the estimated organ center $\mathbf{x}$ is spatially. Geometry of abdominal organs varies due to a) inter-subject variation, b) post-surgical organ shifts, c) postures and d) pathology. To account for the first two factors, the normal spatial models of organ centers are constructed separately for cases with normal anatomy and with different patterns of missing organs due to varying surgical resection procedures. Let MO and NA denote sets of training samples with and without missing organs, respectively. And $MO_{t=1,...,T}$, denotes training samples for the t-th surgical organ resection procedure where T indicates the total number of resection procedures considered and $MO = \bigcup_t MO_t$. Then normal anatomy model M^{na} and missing organ model M^{mo} are defined by the following sets of normal distributions,

$$M^{na} = \{M_l^{na}\} = \{\mathcal{N}(\mathbf{x}; \mathbf{u}_l^{na}, \mathbf{\Sigma}_l^{na})|l = 1, .., L\} \tag{1}$$

$$M^{mo} = \{M_{tl}^{mo}\} = \{\mathcal{N}(\mathbf{x}; \mathbf{u}_l^{mo_t}, \mathbf{\Sigma}_l^{mo_t})|t = 1, .., T, l = 1, .., L\} \tag{2}$$

where $(\mathbf{u}_l^{na}, \mathbf{\Sigma}_l^{na})$ denote the mean and covariance of the center location for organ l averaged over NA, while $(\mathbf{u}_l^{mo_t}, \mathbf{\Sigma}_l^{mo_t})$ denote those averaged over MO_t for the t-th resection procedure.

We define F_l given M^{na} and M^{mo} as follows,

$$\begin{aligned} F_l &= 1 - p(\mathbf{x}|\theta_l) \\ &= min(1 - \mathcal{N}(\mathbf{x}; \mathbf{u}_l^{na}, \mathbf{\Sigma}_l^{na}), \{1 - \mathcal{N}(\mathbf{x}; \mathbf{u}_l^{mo_t}, \mathbf{\Sigma}_l^{mo_t})\}_{t=1,...,T}) \\ &= 1 - max(\mathcal{N}(\mathbf{x}; \mathbf{u}_l^{na}, \mathbf{\Sigma}_l^{na}), \{\mathcal{N}(\mathbf{x}; \mathbf{u}_l^{mo_t}, \mathbf{\Sigma}_l^{mo_t})\}_{t=1,...,T}) \end{aligned} \tag{3}$$

where $\theta_l = ((\mathbf{u}_l^{na}, \mathbf{\Sigma}_l^{na}), \{(\mathbf{u}_l^{mo_t}, \mathbf{\Sigma}_l^{mo_t})\})$. This measure yields high value when the estimated organ center does not follow trends captured in none of the known normal anatomy or surgical procedure-specific models.

The second measure G_l examines the abnormality in texture pattern homogeneity. For each organ l, a binary mask $B_l(\mathbf{x})$ representing an average shape of the organ is derived from the probabilistic atlas by setting $B_l(\mathbf{x}) = 1$, $\forall \mathbf{x}$ $A_l(\mathbf{x}) = 1$ and zero otherwise. Using these binary masks, intensity entropy E_{lm} $= -\sum_{i=1}^{B} p_{lm}(i) \log p_{lm}(i)$ are computed for each organ l in all training samples of NA, where $p_{lm}(i)$ is a B-bin normalized histogram of intensity values sampled under $B_l(\mathbf{x})$ in the m-th sample. For each organ l, the mean and standard deviation of the entropy distribution $(E_l^{na}, \sigma_l^{na})$ are computed over $\{E_{lm}\}$, forming a normal model of organ-specific texture homogeneities. To evaluate an organ l, the entropy E_l of the organ is computed by overlaying $B_l(\mathbf{x})$ by aligning its gravity center to the estimated organ center in the new image and sampling intensity values within the mask. Then G_l is defined as an abnormality measure of E_l with respect to the normal model,

$$G_l = 1 - p(E_l|\phi_l) = 1 - \mathcal{N}(E_l; E_l^{na}, \sigma_l^{na}) \tag{4}$$

where $\phi_l = (E_l^{na}, \sigma_l^{na})$.

The third measure H_l is defined as a linear combination of F_l and G_l,

$$H_l = \beta F_l + (1 - \beta)G_l \tag{5}$$

where $\beta \in [0, 1]$. Finally, missing organs are detected by applying a threshold function to these measures derived for each organ in a new image for arbitrary number of missing organs per case.

2.3 Multi-Organ Segmentation (MOS) with Missing Organs

As a final step, the base MOS method described in Sec 2.1 can be adopted to missing organ cases by discarding the atlas A_l and the spatio-intensity model $\mathcal{N}(\mathbf{x}; \mathbf{u}_{\mathbf{V}l}, \mathbf{\Sigma}_{\mathbf{V}l})$ corresponding to missing organs during the model fitting and inference procedures. The entire MOS procedure thus consists of three successive steps: 1) the base MOS, 2) MOD with F_l, G_l, or H_l, and 3) the modified MOS without A_l, $\mathbf{u}_{\mathbf{V}l}$, and $\mathbf{\Sigma}_{\mathbf{V}l}$ for the detected missing organs.

3 Experiments

3.1 Data

A total of 44 abdominal CT scans are used in this study. Ten non-contrast thin-slice (1mm) abdominal CT scans of healthy volunteers ($K = 10$) are manually segmented by expert radiologists and used to construct the probability atlas by . The NA set contains 25 contrast-enhanced abdominal CT scans with normal anatomy, while the MO set consists of 9 diseased scans with three types ($t = 1, 2, 3$) of surgical organ removal: i) 5 splenectomy cases (spleen removed), ii) 3 nephrectomy cases (right kidney removed), and iii) 1 splenectomy and

(a)   (b)
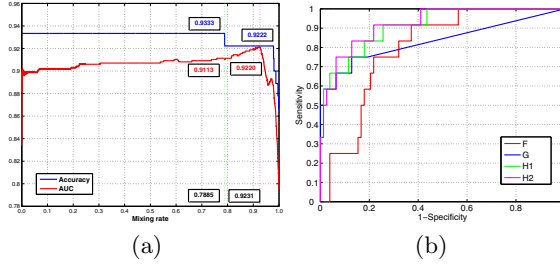
Fig. 2. Quantitative validation of the proposed MOD. (a) Maximum accuracy and AUC values with various mixing rate β for computing the H_l measure. Green and magenta dotted-lines denote β values that yield the maximum accuracy and the maximum AUC, respectively. (b) ROC analysis of MOD with four different measures: red, F_l, blue, G_l, green, $H1$ with $\beta = 0.789$, and pink, $H2$ with $\beta = 0.923$.

nephrectomy case (spleen and left kidney removed). Each scan consists of $512 \times 512 \times 50$ voxel slices with 5mm slice thickness stored in Mayo analyze format. CT scanners from various manufacturers are used to acquire this dataset with the ISOVUE 300 contrast agent. Ten abdominal organs ($L = 10$) are considered in this study: aorta (AO), gallbladder (GB), left/right adrenal glands (LA,RA), liver (LV), left/right kidney (LK,RK), pancreas (PN), spleen (SP), and stomach (ST). For validation, segmentation ground-truth is generated for 9 NA and 9 MO cases by expert researchers with ITK-Snap tool. Fig.1(b,c) illustrate some examples.

3.2   Results

Leave-one-out cross validation is performed to validate the performance of the proposed MOD method on the MO set. For each of the three measures, we evaluated $50,000$ different detection thresholds with a fixed interval between 0 and 1 and derived the receiver operating characteristic (ROC) curves. Maximum accuracies (TP+TN/TP+ TN+FP+FN) with minimum false positive rate was 0.867 and 0.933 for F_l and G_l, respectively. The number of 80 bins (B=80) was used to derive G_l. For H_l, we evaluated $50,000$ different mixing rate β values with a fixed interval between 0 and 1. Fig. 2(a) shows the maximum accuracy and the area under the ROC curve (AUC) computed for various β values. The linear combination did not increase the accuracy measure; the maximum accuracy of 0.933 with highest AUC of 0.911 was found at $\beta = 0.789$ (referred as $H1$). On the other hands, the overall maximum of AUC with 0.922 was found at $\beta = 0.923$ with slightly decreased accuracy of 0.922 (referred as $H2$). Fig. 2(b) shows the ROC curves for F_l, G_l, $H1$, and $H2$, clearly demonstrating the advantage of the proposed linear combination measure. AUC values for F_l and G_l were 0.795 and 0.834, respectively.
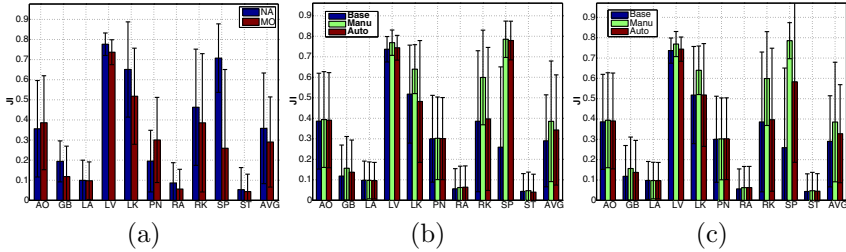
Fig. 3. Average Jaccard index computed for 10 abdominal organs, comparing different MOS methods and datasets. (a) Performance by the base MOS method (*Base*) for normal anatomy (*NA*) and missing organs (*MO*) cases. (b) Comparison of the base and the proposed methods with automatic (*Auto*) and manual (*Manu*) MOD with $\beta = 0.923$ (*H2*) on *MO*. (c) With $\beta = 0.789$ (*H1*).

We next evaluate the proposed MOS method with the missing organ cases. Fig. 3(a) shows organ-wise segmentation accuracy of the base MOS method [4] in Jaccard index (*JI*) on the nine normal anatomy *NA* and the nine diseased *MO* cases as baseline. Liver, left kidney, and spleen have relatively high accuracy. Segmentation of adrenal glands and gall bladder is challenging because they are very small and their shape varies widely. Stomach also yields very low *JI* because its shape and intensity is extremely variant. For most organs, the accuracy for *MO* cases is lower than that for *NA*. The accuracy for spleen and left kidney in *MO* is largely lowered due to missing them in some cases of *MO*. Not only missing organ itself but even neighboring organ, liver, is influenced by right kidney missing such that the bottom of liver is segmented as right kidney that causes the lower accuracy of MO liver.

Fig. 3(b) and (c) compare the accuracy in *JI* for the base and the proposed MOS methods with automatic and manual MOD on *MO* cases with the two versions of H_l measures with $\beta = 0.789$ (*H1*) and $\beta = 0.923$ (*H2*), respectively. The manual MOD specifies which organs are missing according to the ground-truth labels. In both versions, the MOS with manual MOD (*Manu*) performed better than the base method (*Base*), demonstrating the proof-of-concept of our approach in improving segmentation accuracy by explicitly considering missing organs. Our proposed fully-automatic method (*Auto*) outperformed *Base* on average for both versions, although accuracy was lowered from that of *Manu* due to the MOD errors. For spleen, the proposed *Auto* method largely improved *Base* in both versions. The accuracy for the left kidney was slightly improved with $\beta = 0.789$, and that for liver and right kidney was also slightly improved by the both versions of *Auto*.

Fig. 4 shows four illustrative examples for segmenting splenectomy cases (missing spleen). In these examples, spleen (blue), as well as other organs such as gallbladder (dark green) and pancreas (yellow), are fully or partially resected surgically. The examples show that the missing organs are correctly detected by our method and existing neighboring organs, such as left kidney (magenta), is also correctly segmented despite its post-surgical organ shifts. Fig. 5 compares

(a) Missing SP   (b) Missing SP & GB

(c) Missing SP & Partial PN   (d) Missing SP & Partial PN
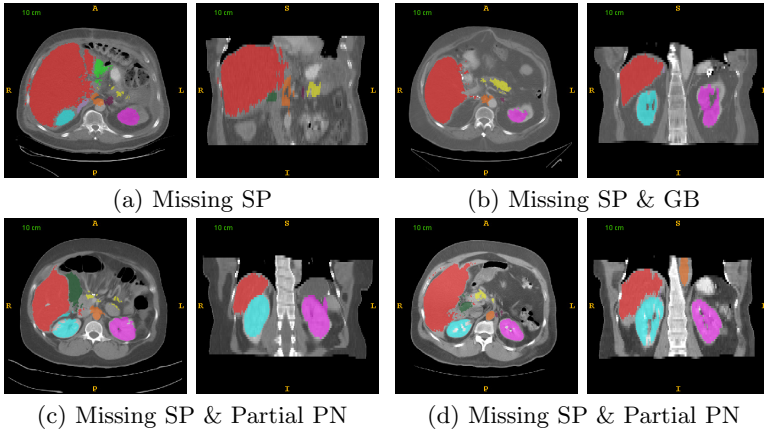
Fig. 4. Four illustrative splenectomy examples of MOS by the proposed *Auto* method. Spleen (blue) is missing in these examples.

(a) *Base*   (b) *Proposed (Auto)*
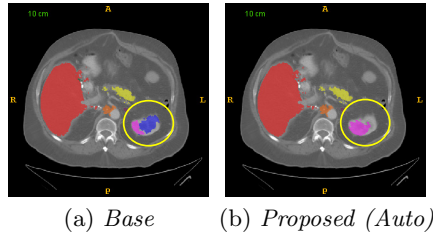
Fig. 5. Segmentation comparison for neighboring organ; (a) the missing spleen (blue) is incorrectly placed inside the left kidney; (b) the improved segmentation

the segmentation results by the base and proposed methods in the splenectomy example in Fig. 4(b). The base method without MOD falsely segments a large part of left kidney (magenta) as (missing) spleen (blue) as shown in Fig. 5(a). Fig. 5(b) clearly shows that the correct MOD of spleen leads to much better segmentation of the neighboring kidney.

4   Conclusions and Discussion

This paper presented novel methods for automatic MOD and atlas-guided MOS that handle missing organs. Our experimental results are promising in that 1) high accuracy of MOD was observed even with the limited number of missing organ cases used in training and 2) the proposed MOS improved the average *JI* accuracy, demonstrating the advantage of our MOD-MOS approach. As our future work, more missing organ cases and surgical resection procedures must be included to further our study in 1) post-surgical organ shifts in finer details and 2) MOD and MOS of partially resected organs that were not addressed in this paper. Finally, we plan to improve the accuracy of our MOS solution, especially

for those difficult organs, by improving our atlas and GMM models, as well as by refining the discontinuous segmentation results by using our results to initialize other graph-based/contour-based segmentation solutions.

References

1. Kobatake, H.: Future CAD in multi-dimensional medical images - project on multi-organ, multi-disease CAD system. Computerized Medical Imaging and Graphics 31, 258–266 (2007)
2. Park, H., Bland, P.H., Meyer, C.R.: Construction of an abdominal probabilistic atlas and its application in segmentation. IEEE Trans. Medical Imaging 22, 483–492 (2003)
3. Zhou, Y., Bai, J.: Multiple abdominal organ segmentation: An atlas-based fuzzy connectedness approach. IEEE Trans. Info. Tech. in Biomed. 11, 348–352 (2007)
4. Shimizu, A., Ohno, R., Ikegami, T., Kobatake, H., Nawano, S., Smutek, D.: Segmentation of multiple organs in non-contrast 3D abdominal CT images. Int. J. CARS 2, 135–143 (2007)
5. Okada, T., Yokota, K., Hori, M., Nakamoto, M., Nakamura, H., Sato, Y.: Construction of Hierarchical Multi-Organ Statistical Atlases and Their Application to Multi-Organ Segmentation from CT Images. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 502–509. Springer, Heidelberg (2008)
6. Seifert, S., Barbu, A., Zhou, S.K., Liu, D., Feulner, J., Huber, M., Suehling, M., Cavallaro, A., Comaniciu, D.: Hierarchical parsing and semantic navigation of full body CT data. In: Proc. SPIE Conf. Medical Imaging (2008)
7. Yao, J., Summers, R.M.: Statistical Location Model for Abdominal Organ Localization. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009, Part II. LNCS, vol. 5762, pp. 9–17. Springer, Heidelberg (2009)
8. Linguraru, M.G., Pura, J.A., Chowdhury, A.S., Summers, R.M.: Multi-organ Segmentation from Multi-phase Abdominal CT via 4D Graphs Using Enhancement, Shape and Location Optimization. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part III. LNCS, vol. 6363, pp. 89–96. Springer, Heidelberg (2010)
9. Linguraru, M.G., Sandberg, J.K., Li, Z., Pura, J.A., Summers, R.M.: Atlas-based automated segmentation of spleen and liver using adaptive enhancement estimation. Medical Physics 37, 771–783 (2010)
10. Liu, X., Linguraru, M.G., Yao, J., Summers, R.M.: Organ Pose Distribution Model and an MAP Framework for Automated Abdominal Multi-organ Localization. In: Liao, H., Edwards, P.J., Pan, X., Fan, Y., Yang, G.-Z. (eds.) MIAR 2010. LNCS, vol. 6326, pp. 393–402. Springer, Heidelberg (2010)
11. Guld, M.O., Kohnen, M., Keysers, D., Schubert, H., Wein, B.B., Bredno, J., Lehmann, T.M.: Quality of DICOM header information for image categorization. In: Proc. SPIE Conf. Medical Imaging, vol. 4685, pp. 280–287 (2002)
12. Rueckert, D., Sonoda, L.I., Hayes, C., Hill, D.L.G., Leach, M.O., Hawkes, D.J.: Non-rigid registration using free-form deformations: Application to breast MR images. IEEE Trans. Medical Imaging 18, 712–721 (1999)
13. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. J. Roy Stats. Soc. Series B 39, 1–38 (1977)

Non-local STAPLE:
An Intensity-Driven Multi-atlas Rater Model

Andrew J. Asman and Bennett A. Landman

Electrical Engineering, Vanderbilt University, Nashville, TN, USA 37235
{andrew.j.asman,bennett.landman}@vanderbilt.edu

Abstract. Multi-atlas segmentation provides a general purpose, fully automated class of techniques for transferring spatial information from an existing dataset ("atlases") to a previously unseen context ("target") through image registration. The method used to combine information after registration ("label fusion") has a substantial impact on the overall accuracy and robustness. In practice, weighted voting techniques have dramatically outperformed algorithms based on statistical fusion (i.e., algorithms that incorporate rater performance into the estimation process — STAPLE). We posit that a critical limitation of statistical techniques (as generally proposed) is that they fail to incorporate intensity seamlessly into the estimation process and models of observation error. Herein, we propose a novel statistical fusion algorithm, Non-Local STAPLE, which merges the STAPLE framework with a non-local means perspective. Non-Local STAPLE (1) seamlessly integrates intensity into the estimation process, (2) provides a theoretically consistent model of multi-atlas observation error, and (3) largely bypasses the need for group-wise unbiased registrations. We demonstrate significant improvements in two empirical multi-atlas experiments.

Keywords: Simultaneous Truth And Performance Level Estimation (STAPLE), Statistical Label Fusion, Rater Models, Multi-Atlas Segmentation.

1   Introduction

The *de facto* standard baseline for large-scale, consistent, and robust segmentation is to perform a multi-atlas segmentation in which a collection of canonical atlases (with labels) are registered to a target-of-interest [1, 2]. Here, we focus on the problem of resolving voxelwise conflicts between the registered atlases (i.e., "label fusion").

Voting fusion strategies (e.g., a majority vote) have long provided robust segmentations. Recently, weighted voting using global [3], local [4], semi-local [5] and non-local [6] intensity similarities between the atlases and the target have demonstrated significant improvements in segmentation accuracy. Particularly for neurological applications, highly local weights have provided the most consistent and accurate segmentation estimates [4, 5].

In contrast to voting, statistical fusion strategies (e.g., Simultaneous Truth And Performance Level Estimation, STAPLE [7]) directly integrate a model of rater behavior (i.e., labeling error probabilities). Despite elegant theory and success with
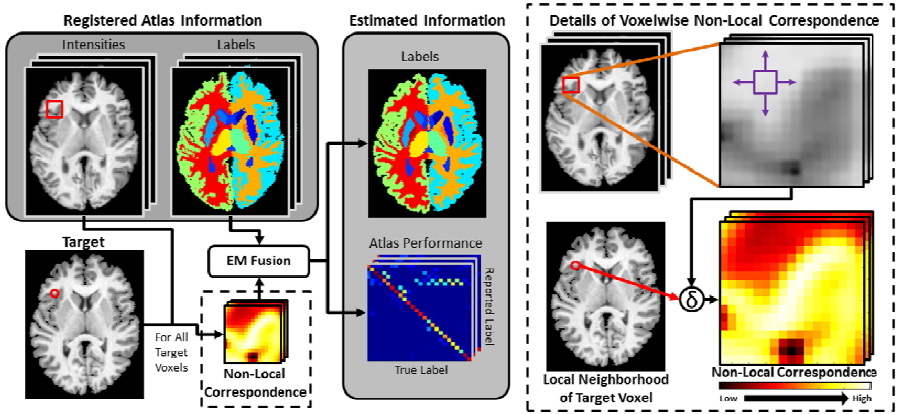
Fig. 1. Flowchart of the Non-Local STAPLE (NLS) algorithm. NLS uses the atlas and target intensities to construct a non-local correspondence and integrates this into the estimation process. Point-wise correspondence is constructed in a traditional non-local means approach.

human raters, applications to the multi-atlas context have proven problematic [5, 8, 9]. In response, a myriad of variations have been proposed to account for spatially varying difficulty [10] and performance [8, 9]. Nevertheless, seamless integration of exogenous intensity information into the STAPLE context has proven difficult — efforts have relied upon ignoring voxels based on intensity similarities [9]. As a result, statistical fusion strategies are often less accurate in clinical multi-atlas applications.

Herein, we propose a novel statistical fusion algorithm (Non-Local STAPLE — NLS) that merges the STAPLE framework with a non-local means perspective [11]. NLS models the registered atlases as collections of volumetric patches containing both intensity and label information and uses the non-local criteria [6, 11] to resolve imperfect correspondence (Figure 1). Through this reformulation, we seamlessly integrate exogenous intensity information into the estimation process to provide a theoretically consistent model of multi-atlas observation error. We derive the theoretical basis governing NLS, demonstrate significant improvement over premier fusion algorithms on two distinct datasets (CT thyroid and MR whole-brain segmentation), and assess the sensitivity of NLS to the various model parameters.

2   Theory

Consider a target gray-level image represented as a vector, $I \in \mathbb{R}^{N \times 1}$. Let $T \in L^{N \times 1}$ be the latent representation of the true target segmentation, where $L = \{0, \dots, L-1\}$ is the set of possible labels. Consider a collection of R registered atlases with associated intensity values, $A \in \mathbb{R}^{N \times R}$, and label decisions, $D \in L^{N \times R}$. Let $\theta \in \mathbb{R}^{R \times L \times L}$ parameterize the raters (registered atlases) performance level. Each element of θ, $\theta_{js's}$, represents the probability that rater j observes label s' given that the true label is s at a given target voxel and the *corresponding* voxel on the associated atlas

— i.e., $\theta_{js's} \equiv p(D_{i^*j} = s', \mathbf{A}|T_i = s, \mathbf{I})$, where i^* is the voxel on atlas j that corresponds to target voxel i. Throughout, the index variables i, i^* and i' will be used to iterate over the voxels, s and s' over the labels, and j over the registered atlases.

2.1   The Non-local STAPLE Algorithm

NLS uses an Expectation-Maximization (EM) approach to estimate the true latent segmentation based on the target intensities, atlas information, and the rater performance level parameters (Figure 1). Estimation of the true segmentation (E-Step) follows [7]. Let $\mathbf{W} \in \mathbb{R}^{L \times N}$, where W_{si} represents the probability that the true label associated with voxel i is label s. Using a Bayesian expansion and conditional independence between the raters, the solution for $\mathbf{W}$ on iteration k is

$$
\begin{aligned}
W_{si}^{(k)} &\equiv f(T_i = s|\mathbf{D}, \mathbf{A}, \mathbf{I}, \boldsymbol{\theta}^{(k)}) \\
&= \frac{f(T_i = s) \prod_j f(D_{i^*j} = s', \mathbf{A}_j|T_i = s, I_i, \theta_{js's}^{(k)})}{\sum_n f(T_i = n) \prod_j f(D_{i^*j} = s', \mathbf{A}_j|T_i = n, I_i, \theta_{js'n}^{(k)})}
\end{aligned}
\tag{1}
$$

where $f(T_i = s)$ is a voxelwise *a priori* distribution of the underlying segmentation, and D_{i^*j} is the label decision by atlas j and corresponding voxel i^*.

In NLS, we assume that we do not know which voxel i^* on atlas j corresponds with voxel i on the target. We approximate the expansion with the expected value of Eq. 1 based on the probability of correspondence across the images and an assumption of conditional independence between the labels and intensity:

$$
\begin{aligned}
f(D_{i^*j} = s', \mathbf{A}_j|T_i = s, I_i, \theta_{js's}^{(k)}) &\approx E[f(\mathbf{D}_j, \mathbf{A}_j|T_i = s, I_i, \boldsymbol{\theta}_{js}^{(k)})] \\
&= E[f(\mathbf{D}_j|T_i = s, \boldsymbol{\theta}_{js}^{(k)})f(\mathbf{A}_j|I_i)] \\
&= \sum_{i' \in \mathcal{N}_s(i)} f(D_{i'j} = s'|T_i = s, \theta_{js's}^{(k)}) f(A_{i'j}|I_i)
\end{aligned}
\tag{2}
$$

where $\mathcal{N}_s(i)$ is the *search neighborhood* of voxel i, and $f(A_{i'j}|I_i)$ is the probability of the non-local correspondence between the target at voxel i and voxel i' on atlas j. We use a standard non-local means approach to define:

$$
f(A_{i'j}|I_i) \equiv \alpha_{ji'i} = \frac{1}{Z_\alpha} \exp\left(-\frac{\|\mathcal{N}_p(A_{i'j}) - \mathcal{N}_p(I_i)\|_2^2}{2\sigma_i^2} - \frac{d_{ii'}^2}{2\sigma_d^2}\right)
\tag{3}
$$

where $\mathcal{N}_p(\cdot)$ is the set of intensities in the *patch neighborhood* of a given intensity location, $d_{ii'}$ is the Euclidean distance between voxels i and i' in image space, and σ_i and σ_d are the standard deviations of the assumed Gaussian distributions governing the intensity similarity and the Euclidean distance-based decay, respectively. Lastly, Z_α is a partition function that enforces the constraint that $\sum_{i' \in \mathcal{N}_s(i)} f(A_{i'j}|I_i) = 1$.

Finally, we revisit Eq. 2 and, using the fact that $s' = D_{i'j}$, find that the final representation for the voxelwise label probabilities is

$$W_{si}^{(k)} = \frac{f(T_i = s) \prod_j \sum_{i' \in \mathcal{N}_s(i)} \theta_{js's}^k \alpha_{ji'i}}{\sum_n f(T_i = n) \prod_j \sum_{i' \in \mathcal{N}_s(i)} \theta_{js'n}^k \alpha_{ji'i}}. \tag{4}$$

The estimate of the performance level parameters (M-Step) is obtained by finding the parameters that maximize the expected value of the conditional log likelihood function found in Eq. 4.

$$\begin{aligned}
\boldsymbol{\theta}_j^{(k+1)} &= \arg\max_{\boldsymbol{\theta}_j} \sum_i E\left[\ln f(D_{i^*j} = s', \boldsymbol{A}_j | T_i = s, I_i, \theta_{js's}^{(k)}) \,\middle|\, \boldsymbol{D}, \boldsymbol{A}, \boldsymbol{I}, \boldsymbol{\theta}^{(k)}\right] \\
&= \arg\max_{\boldsymbol{\theta}_j} \sum_i \sum_s W_{si}^{(k)} \ln\left(f(D_{i^*j} = s', \boldsymbol{A}_j | T_i = s, I_i, \theta_{js's}^{(k)})\right).
\end{aligned} \tag{5}$$

Noting the constraint that each row of the rater performance matrix must sum to one to be a valid probability mass function (i.e., $\sum_{s'} \theta_{js's}^{(k)} = 1$), we can maximize the performance level parameters by formulating the constrained optimization problem using a LaGrange multiplier. After taking the element-wise partial derivative and using the constraint that $\sum_{i' \in \mathcal{N}_s(i)} f(A_{i'j}|I_i) = 1$, the performance update becomes

$$\theta_{js's}^{(k+1)} = \frac{\sum_i \left(\sum_{i' \in \mathcal{N}_s(i): D_{i'j}=s'} \alpha_{ji'i}\right) W_{si}^{(k)}}{\sum_i W_{si}^{(k)}}. \tag{6}$$

2.2   Initialization and Convergence

As is typical [7], NLS was initialized with performance parameters equal to 0.95 along the diagonal and randomly setting the off-diagonal elements to fulfill the required constraints. For all presented experiments, the voxelwise label prior, $f(T_i = s)$, was initialized using the probabilities from a "weak" log-odds majority vote (i.e., decay coefficient set to 0.5) [5], and the search neighborhood, $\mathcal{N}_s(i)$, was initialized to an $11 \times 11 \times 11$ window centered at the target voxel of interest. Several values for the patch neighborhood, $\mathcal{N}_p(\cdot)$, are considered in this manuscript all of which are centered at the voxels of interest. Unless otherwise noted, the values of the standard deviation parameters, σ_i and σ_d, were set to 0.1 and 2, respectively. Lastly, convergence of the algorithm was detected when the average change in the on-diagonal elements of the performance level parameters fell below 10^{-4}.

3   Methods and Results

As benchmarks, we compare to a log-odds majority vote (MV) [5], a locally weighted vote (LWV) [5], and STAPLE [7]. For the voting algorithms, the implementation was the same as suggested in [5]. Note that LWV has a parameter that is essentially equivalent to the NLS parameter σ_i. For fairness of comparison, this parameter was set to the same value (herein, 0.1) for both algorithms. STAPLE was initialized using the same value for $f(T_i = s)$ as NLS. For both STAPLE and NLS, "consensus voxels" (herein, voxels where $\max_s f(T_i = s) > 0.95$) were ignored. For all experiments, the

atlases were intensity normalized to the 25th and 75th percentiles. All pair-wise registrations were performed using an initial affine registration followed by a non-rigid procedure (Adaptive Bases Algorithm [12]). After registration the images were cropped to obtain a reasonable region of interest. Quantitative accuracy was primarily assessed using the Dice Similarity Coefficient (DSC).

3.1   Thyroid Multi-atlas Segmentation

First, we analyzed the fusion accuracy on an empirical multi-atlas approach for thyroid segmentation using a collection of 15 head and neck atlases. The computed tomography (CT) images used in this experiment were collected from consenting patients who underwent intensity-modulated radiation therapy. The patients were injected with 80mL of Optiray 320, a 68% iversol-based nonionic contrast agent. Each image has a voxel size of $1 \times 1 \times 3$ mm^3. We performed a leave-one-out cross-validation experiment (i.e., 14 atlases per segmentation estimate) to assess fusion accuracy. NLS was run using various patch neighborhood sizes ($1 \times 1 \times 1$, $3 \times 3 \times 3$, $5 \times 5 \times 3$, and $7 \times 7 \times 3$).

NLS substantially improved thyroid segmentation accuracy with the $3 \times 3 \times 3$ patch neighborhood significantly outperforming all other algorithms (p < .05, Figure 2A). Median DSC performance was improved by 0.05 over LWV and 0.08 over STAPLE. The quantitative results seen in Figure 2A show the accuracy (in terms of the DSC) of the considered algorithms across the 15 atlases. Note the significant outliers in the results for the voting-based algorithms. Qualitative results can be seen
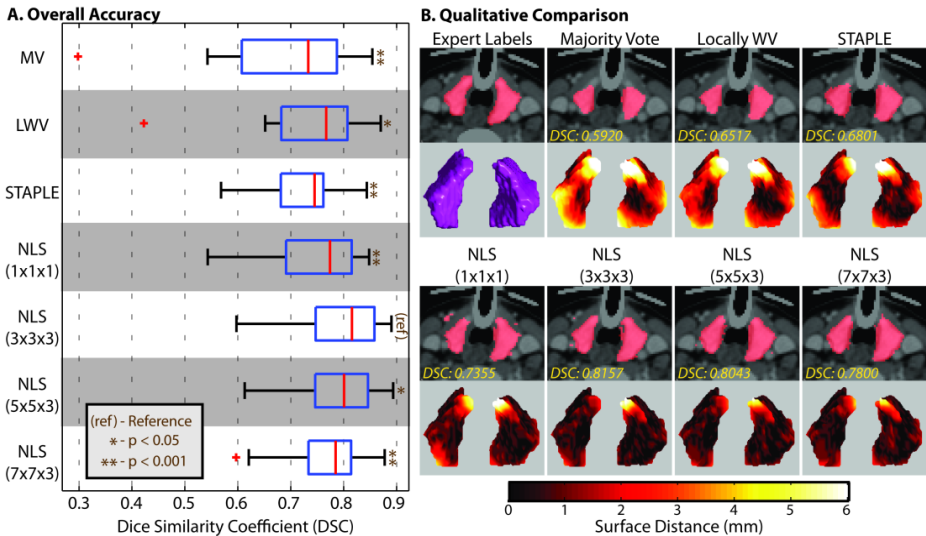
Fig. 2. Results of the empirical multi-atlas segmentation of the thyroid. The quantitative results (A) show that NLS provides significant improvement, with a $3 \times 3 \times 3$ patch neighborhood significantly outperforming all other algorithms. The qualitative results (B) demonstrate that NLS provides improvement in terms of shape, boundary and point-wise surface distance error.

in Figure 2B, where, for all considered algorithms, a representative slice and a 3D rendering of the point-wise surface distance error is presented. The various estimations from NLS are all qualitatively superior to the other benchmarks, as they more accurately estimated the underlying shape and size and resulted in substantial reductions in point-wise surface distance error. For small patch neighborhoods (e.g., $1 \times 1 \times 1$), it is evident that high quality boundaries are estimated but "speckle noise" is more likely to be apparent. Alternatively, for larger windows, estimations are smoother but sacrifice the high quality boundary estimation.

3.2   Whole-Brain Multi-atlas Segmentation

Second, we examine fusion accuracy on an empirical experiment for whole-brain segmentation. A collection of 15 brains (OASIS, www.oasis-brains.org) were manually labeled (www.braincolor.org) by an expert anatomist. For each atlas a collection 26 labels were considered ranging from large structures (e.g. cortical gray matter) to smaller deep brain structures. All images were 1mm isotropic. To assess overall accuracy, we performed a cross-validation experiment using 5 to 14 atlases per target atlas. The per-label accuracy was assessed using 5 atlases per target. Lastly, the sensitivity of NLS with respect to the parameters σ_i and σ_d was assessed. Due to the large number of labels and limited atlases, STAPLE results were poor (not shown).
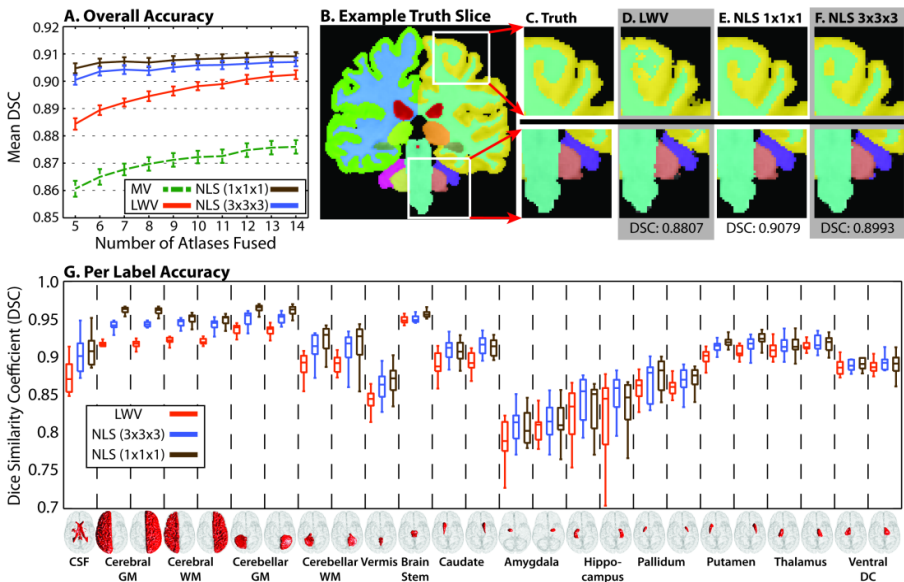
Fig. 3. Results from the empirical whole-brain experiment. Overall (A) and qualitative (B-F) results show NLS (with a single voxel neighborhood) significantly outperforming the other algorithms. Per-label results (G) show consistent improvement regardless of label size or location.

The results of the cross-validation (Figure 3) demonstrate that NLS provides consistent improvement in segmentation accuracy. For overall accuracy (reported as mean DSC, Figure 3A), NLS resulted in significant improvement ($p < 0.05$) over the other algorithms regardless of the number of atlases and provided estimates that are less dependent upon the number of atlases fused. Unlike the thyroid results, a single voxel neighborhood resulted in consistent improvement over larger neighborhood sizes. NLS using a single voxel neighborhood resulted in qualitatively more accurate segmentations (Figure 3B-3F). The per-label results (Figure 3G) demonstrate that, particularly for the larger labels, the NLS estimates are vastly superior to a locally weighted vote. NLS using a $1 \times 1 \times 1$ patch neighborhood resulted in significantly superior ($p < 0.05$) results over LWV on 23 out of 26 labels and for 16 out of 26 labels over NLS using a $3 \times 3 \times 3$ patch neighborhood. Neither MV nor LWV was significantly superior to either NLS approach for any label.
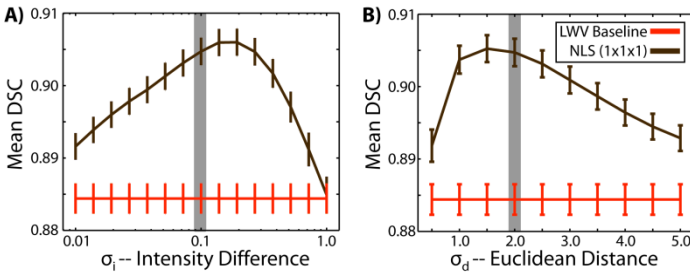
Fig. 4. Sensitivity to NLS model parameters. The sensitivity of NLS to σ_i (A) and σ_d (B) demonstrate degraded performance for values that are either too small or too large. Regardless, consistent improvement over a locally weighted vote is achieved. Gray outlines indicate the values used in the previously presented experiments.

Lastly, the sensitivity of NLS to two of the model parameters, σ_i and σ_d (Eq. 3), can be appreciated in Figure 4. NLS accuracy decreases for σ_i values that result in segmentations that are overly noisy (small σ_i values) or overly smooth (larger σ_i values) (Figure 4A). Note that the value of this parameter is largely dependent upon the intensity normalization process (i.e., the relative distribution of atlas and target intensities). NLS sensitivity to σ_d, which can be interpreted as a proxy for sensitivity to search window size, is presented in Figure 4B. For values that are too small, accuracy sharply diminishes as too few voxels are used in constructing the non-local correspondence. Alternatively, values that are too large result in the inclusion of regions of the image that are not anatomically indicative of the label of interest.

4   Discussion

Non-Local STAPLE represents the first statistical fusion algorithm that (1) creates a cohesive theoretical model specifically targeting registered atlas observation behavior, and (2) incorporates intensity seamlessly into the core of the fusion framework's rater model. Both of these goals are accomplished through the reformulation of the

STAPLE algorithm from a non-local means perspective and the integration of the concept of non-local correspondence into the estimation process. NLS models atlas observation behavior by learning which label would have been observed, given perfect correspondence between the target and the atlases. NLS overcomes several of the current obstacles that plague both the accuracy and theoretical underpinning of label fusion algorithms. We demonstrated superior performance over premier label fusion algorithms on two empirical multi-atlas experiments for segmentation of the thyroid (Figure 2) and whole-brain segmentation (Figures 3 and 4).

While the sensitivity of NLS is assessed with respect to the number of atlases fused (Figure 3A) and the parameters σ_i and σ_d (Figure 4), several questions still persist in order to understand the optimality of the algorithm. For example, the effect of using an alternative similarity metric (e.g., normalized correlation coefficients or mutual information) to the assumed Gaussian difference model presented here (Eq. 3) need to be investigated. Alternative similarity measures may dramatically lessen the impact of noise in the intensity images and the need for accurate intensity normalization between the target and the atlases. Additionally, automated techniques for determining optimal window sizes (i.e. $\mathcal{N}_p$ and $\mathcal{N}_s$) and initialization strategies would provide valuable advancements for the applicability of NLS to new problem spaces.

Recently, several advancements to the STAPLE framework have been suggested to account for spatially varying labeling difficulty [2, 10] and rater performance [8, 9]. We propose that these advancements could be integrated in a straightforward manner into the current theoretical formulation governing NLS. Further investigation into their applicability to the NLS framework represents fascinating areas of continuing research. Additionally, while not presented here, future investigation into the relationship between NLS and non-local voting based algorithms [6] is critical to understanding the importance of integrating performance level estimates into the multi-atlas estimation process. Lastly, integration of Markov Random Fields (MRF) [5, 7] and global/local atlas pre-selection [9] could provide valuable benefits in terms of segmentation accuracy (e.g., limiting the effects seen in Figure 2B).

References

1. Heckemann, R.A., et al.: Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. Neuroimage 33, 115–126 (2006)
2. Rohlfing, T., et al.: Performance-based classifier combination in atlas-based image segmentation using expectation-maximization parameter estimation. IEEE Trans. Med. Imaging 23, 983–994 (2004)
3. Artaechevarria, X., et al.: Combination strategies in multi-atlas image segmentation: Application to brain MR data. IEEE Trans. Med. Imaging 28, 1266–1277 (2009)
4. Isgum, I., et al.: Multi-atlas-based segmentation with local decision fusion—Application to cardiac and aortic segmentation in CT scans. IEEE Trans. Med. Imaging 28, 1000–1010 (2009)

5. Sabuncu, M.R., et al.: A generative model for image segmentation based on label fusion. IEEE Trans. Med. Imaging 29, 1714–1729 (2010)
6. Coupé, P., et al.: Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation. Neuroimage 54, 940–954 (2011)
7. Warfield, S.K., et al.: Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. IEEE Trans. Med. Imaging 23, 903–921 (2004)
8. Asman, A.J., Landman, B.A.: Characterizing Spatially Varying Performance to Improve Multi-atlas Multi-label Segmentation. In: Székely, G., Hahn, H.K. (eds.) IPMI 2011. LNCS, vol. 6801, pp. 85–96. Springer, Heidelberg (2011)
9. Weisenfeld, N.I., Warfield, S.K.: Learning Likelihoods for Labeling (L3): A General Multi-*Classifier* Segmentation Algorithm. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part III. LNCS, vol. 6893, pp. 322–329. Springer, Heidelberg (2011)
10. Asman, A., Landman, B.: Robust Statistical Label Fusion through Consensus Level, Labeler Accuracy and Truth Estimation (COLLATE). IEEE Trans. Med. Imaging 30, 1779–1794 (2011)
11. Buades, A., et al.: A non-local algorithm for image denoising. In: Computer Vision and Pattern Recognition (CVPR), vol. 62, pp. 60–65. IEEE (2005)
12. Rohde, G.K., et al.: The adaptive bases algorithm for intensity-based nonrigid image registration. IEEE Trans. Med. Imaging 22, 1470–1479 (2003)

Shape Prior Modeling Using Sparse Representation and Online Dictionary Learning

Shaoting Zhang[1],[*], Yiqiang Zhan[2],[*],
Yan Zhou[3], Mustafa Uzunbas[2], and Dimitris N. Metaxas[1]

[1] Department of Computer Science, Rutgers University, Piscataway, NJ, USA
[2] CAD R&D, Siemens Healthcare, Malvern, PA, USA
[3] Elekta Inc., Maryland Heights, MO, USA
shaoting@cs.rutgers.edu, yiqiang.zhan@siemens.com

Abstract. The recently proposed Sparse Shape Composition (SSC) opens a new avenue for shape prior modeling. Instead of assuming any parametric model of shape statistics, SSC incorporates shape priors on-the-fly by approximating a shape instance (usually derived from appearance cues) by a sparse combination of shapes in a training repository. Theoretically, one can increase the modeling capability of SSC by including as many training shapes in the repository. However, this strategy confronts two limitations in practice. *First*, since SSC involves an iterative sparse optimization at run-time, the more shape instances contained in the repository, the less run-time efficiency SSC has. Therefore, a compact and informative shape dictionary is preferred to a large shape repository. *Second*, in medical imaging applications, training shapes seldom come in one batch. It is very time consuming and sometimes infeasible to re-construct the shape dictionary every time new training shapes appear. In this paper, we propose an online learning method to address these two limitations. Our method starts from constructing an initial shape dictionary using the K-SVD algorithm. When new training shapes come, instead of re-constructing the dictionary from the ground up, we update the existing one using a block-coordinates descent approach. Using the dynamically updated dictionary, sparse shape composition can be gracefully scaled up to model shape priors from a large number of training shapes without sacrificing run-time efficiency. Our method is validated on lung localization in X-Ray and cardiac segmentation in MRI time series. Compared to the original SSC, it shows comparable performance while being significantly more efficient.

1 Introduction

Sparse Shape Composition (SSC) [11] is a recently proposed method for shape prior modeling. Different from previous methods [3], which often assume a parametric model for shape statistics, SSC is a non-parametric method that approximates an input shape usually derived from low level appearance features, by a sparse combination of other shapes in a repository. In this way, shape priors

---

[*] Corresponding authors.

are incorporated on-the-fly. SSC is able to correct gross errors of input shape and can preserve shape details even if they are not statistically significant in the training repository.

Theoretically, the more shape instances contained in the shape repository, the more shape modeling capacity SSC has. However, a repository including a large number of shapes adversely affects the efficiency of SSC, which iteratively performs sparse optimization at run-time. To reduce the redundancy in the shape repository and the computational cost, one can learn a compact and informative dictionary. Unfortunately, dictionary learning sometimes confronts another limitation. In medical imaging applications, training shape instances seldom come in one batch. If the dictionary needs to be completely re-learned every time when new training shapes come, the learning process will become very time consuming and sometimes infeasible.

In this paper, we propose an online learning method to address these two limitations. Our method starts from learning an initial dictionary offline using available training shapes. The K-SVD method is employed to learn the initial dictionary due to its flexibility and accelerated convergency. When new training shapes come, instead of re-constructing the dictionary from the ground up, we use an online dictionary learning method [7] to update the shape dictionary on-the-fly. With more and more new training shapes, our shape dictionary keeps updated to contain shape priors from all of them. Hence, sparse shape composition performed on the shape dictionary achieves two advantages: 1) The run-time efficiency of the shape composition is not sacrificed given much more training shapes. 2) SSC can be gracefully scaled-up to apply shape priors from, theoretically, infinite number of training shapes.

Relevant Work: Related studies can be traced to two categories, shape modeling and sparse dictionary learning. In the former category, most previous studies [3,5,6,9,10] aim to model shape priors using a parametric model, e.g., multi-variant Gaussian [3] and hierarchical diffusion wavelet [6]. SSC is the first shape modeling method using sparse representation theory. Sparse dictionary learning methods have been extensively studied in signal processing domain. Popular ones include optimal direction (MOD) and K-SVD [1]. While these methods require the access of all training samples, a recently proposed online dictionary learning [7] allows an efficient dictionary update only based on new samples. Although dictionary learning has been successfully applied on low level image processing tasks, to the best of our knowledge, the proposed method is the first one to employ them for high-level shape prior modeling.

2   Methodology

In this section, we will first briefly introduce standard Sparse Shape Composition. Dictionary learning technologies that aims to tackle the two limitations of SSC will be presented afterwards.

Sparse Shape Composition: SSC is designed based on two observations: 1) After being aligned to a common canonical space, any shape can be approx-

imated by a sparse linear combination of other shape instances in the same shape category. Approximation residuals might come from inter-subject variations. 2) If the shape to be approximated is derived by appearance cues, residual errors might include gross errors from detection/segementaion errors. However, such errors are sparse as well. Accordingly, shape priors can be incorporated *on-the-fly* through shape composition, which is formulated as a sparse optimization problem as follows.

In SSC, a shape is represented by a contour (2D) or a triangle mesh (3D) which consists of a set of vertices. Denote the input shape as $\mathbf{v}$, where $\mathbf{v} \in \mathbf{R}^{\mathfrak{D}N}$ is a vector concatenated by coordinates of its N vertices, where $\mathfrak{D} = \{2,3\}$ denotes the dimensionality of the shape modeling problem. (In the remainder of this paper, any shape instance is defined as a vector in the same way.) Assume $D = [\mathbf{d}_1, \mathbf{d}_2, ..., \mathbf{d}_K] \in \mathbf{R}^{\mathfrak{D}N \times K}$ is a large shape repository that includes K accurately annotated and pre-aligned shape instances $\mathbf{d}_i$. The approximation of $\mathbf{v}$ by D is then formulated as an optimization problem:

$$\underset{\mathbf{x}, \mathbf{e}, \beta}{\arg\min} \ \|T(\mathbf{v}, \beta) - D\mathbf{x} - \mathbf{e}\|_2^2 + \lambda_1 \|\mathbf{x}\|_1 + \lambda_2 \|\mathbf{e}\|_1, \qquad (1)$$

where $T(\mathbf{v}, \beta)$ is a global transformation operator with parameter β, which aligns the input shape $\mathbf{v}$ to the common canonical space of D. The key idea of SSC lies in the second and third terms of the objective function. In the second term, the $L1$-norm of $\mathbf{x}$ ensures that the nonzero elements in $\mathbf{x}$, i.e., the linear combination coefficients, is sparse [2]. Hence, only a sparse set of shape instances can be used to approximate the input shape, which prevents the overfitting to errors from missing/misleading appearance cues. In the third term, the same sparse constraint applies on $\mathbf{e} \in \mathbf{R}^{\mathfrak{D}N}$, the large residual errors, which incorporates the observation that gross errors might exist but are occasional. Eq. 1 is optimized using an Expectation-Maximization (EM) style algorithm, which alternatively optimizes β ("E" step) and $\mathbf{x}$, $\mathbf{e}$ ("M" step). "M" step employs a typical convex solver, e.g., interior-point convex solver [8] in this study.

Shape Dictionary Learning: Theoretically, the more shape instances in D, the larger shape modeling capacity SSC has. However, the run-time efficiency of SSC is also determined by the size of the shape repository matrix $D \in \mathbf{R}^{\mathfrak{D}N \times K}$. More specifically, the computational complexity of the interior-point convex optimization solver is $\mathcal{O}(N^2 K)$ per iteration [8], which means the computational cost will increase quickly with the increase of K, the number of the shape instances in the shape repository. Note that $\mathcal{O}(N^2 K)$ is the computational complexity for *one* iteration. Empirically, with larger K, it usually takes more iterations to convergency, which further decreases the algorithm speed.

In fact, owing to the similar shape characteristics across the population, these K shape instances usually contain lots of redundant information. Instead of including all of them, D should only contain "representative" shapes. This is exactly a dictionary learning problem, which has been extensively investigated in signal processing community. More specifically, a well learned dictionary should have a compact set of "atoms" that are able to sparsely approximate other

---

Algorithm 1. Online learn and update dictionary, using mini-batch mode.

Input: Initialized dictionary $D_0 \in \mathbf{R}^{n \times k}$, input data $Y = [y_1, y_2, ..., y_K], y_i \in \mathbf{R}^n$, number of iterations T, regularization parameter $\lambda \in R$.
Output: Learned dictionary D_T.
$A_0 = 0, B_0 = 0$.
for $t = 1 \rightarrow T$ **do**
    Randomly draw a set of $y_{t,1}, y_{t,2}, ..., y_{t,\eta}$.
    for $i = 1 \rightarrow \eta$ **do**
        Sparse coding: $x_{t,i} = \underset{x \in \mathbf{R}^k}{\arg\min} \frac{1}{2}\|y_{t,i} - D_{t-1}x\|_2^2 + \lambda\|x\|_1$.
    end for
    $A_t = \beta A_{t-1} + \sum_{i=1}^{\eta} x_{t,i}x_{t,i}^T, B_t = \beta B_{t-1} + \sum_{i=1}^{\eta} y_{t,i}x_{t,i}^T,$
    where $\beta = \frac{\theta+1-\eta}{\theta+1}$, and $\theta = t\eta$ if $t < \eta$, $\theta = \eta^2 + t - \eta$ otherwise.
    Dictionary update: Compute D_t, so that:
    $\underset{D}{\arg\min} \frac{1}{t} \sum_{i=1}^{t} \frac{1}{2}\|y_i - Dx_i\|_2^2 + \lambda\|x_i\|_1 = \underset{D}{\arg\min} \frac{1}{t}\left(\frac{1}{2}Tr\left(D^T D A_t\right) - Tr(D^T B_t)\right).$
end for

---

signals. In our study, shape dictionary is learned using K-SVD [1], a popular dictionary learning method because of its accelerated converging speed.

Online Shape Dictionary Update: Using the compact dictionary derived by K-SVD, the run-time efficiency of SSC is dramatically improved, as the number of atoms in D is much less than the number of training shapes. However, K-SVD requires all training shapes available in the "dictionary update" step, which can not be satisfied in a lot of medical applications. For example, owing to the expensive cost, manual annotations of anatomical structures often come gradually from different radiologists/technicians. Re-construction of the dictionary D with every batch of new training shapes is very time consuming and not always feasible. To tackle this problem, we employ a recently proposed online dictionary method [7] to update the shape dictionary.

Algorithm 1 shows the framework of online dictionary learning for sparse coding. Starting from an initial dictionary learned by K-SVD, it iteratively employs two stages until converge, sparse coding and dictionary update. Sparse coding aims to find the sparse coefficient x_i for each signal y_i:

$$x_i = \underset{x \in \mathbf{R}^k}{\arg\min} \frac{1}{2}\|y_i - Dx\|_2^2 + \lambda\|x\|_1 \tag{2}$$

where D is the initialized dictionary or dictionary computed from the previous iteration. LARS-Lasso algorithm [4] is employed to solve this step. The dictionary update stage aims to update D based on all discovered $x_i, i \in [1, K]$:

$$\underset{D}{\arg\min} \frac{1}{K} \sum_{i=1}^{K} \frac{1}{2}\|y_i - Dx_i\|_2^2 + \lambda\|x_i\|_1 \tag{3}$$

Based on stochastic approximation, the dictionary is updated efficiently using block-coordinates descent. It is a parameter-free method and does not require
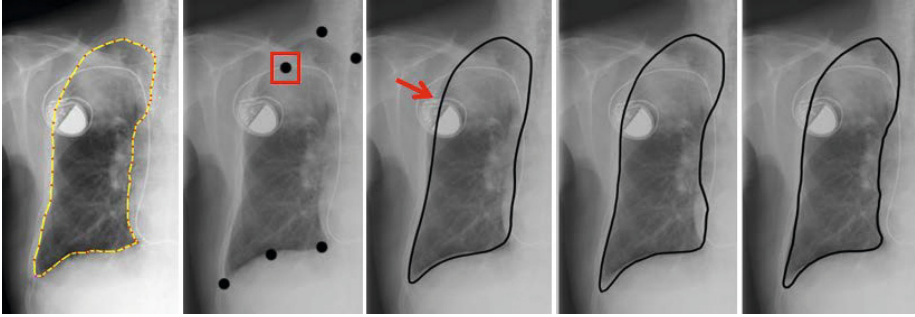
Fig. 1. Comparisons of the localization results. From left to right: manual label, detection results, PCA, SSC, and the online learning based shape refinement results. Due to the erroneous detection (marked by the red box), PCA result moves to the right and is not on the boundary (see the red arrow). Zoom in for better view.

any learning rate tuning. It is important to note that the "dictionary update" step in Algorithm 1 is significantly different from that of K-SVD. Instead of requiring all training shapes, it only exploits a small batch of newly coming data (i.e., $x_i, i \in [1, \eta]$). The dictionary update thereby becomes much faster than K-SVD, as $\eta \ll K$. In this way, we can efficiently update the shape dictionary online by using new data as selected x_i.

Using this online updated dictionary, SSC obtains two additional advantages. 1) The run-time efficiency of shape composition is not sacrificed with much more training shapes. 2) SSC can be gracefully scaled-up to contain shape priors from, theoretically, infinite number of training shapes.

3   Experiments

We validate our algorithm in two applications, lung localization in Chest X-ray, and left ventricle tracking in MRI.

Lung Localization: Chest radiography (X-ray) is a widely used medical imaging modality because of the fast imaging speed and low cost. Localization of lungs in chest radiography not only provides lung shapes, which are critical clues for pathology detection, but also paves the way for other medical image analysis tasks, e.g., cardiac measurements. On one hand, owing to the relatively cheap cost of manual/semi-automatic annotations of lungs in X-ray images, it is possible to get a large number of lung shapes for training. On the other hand, however, training lung shapes seldom come in one batch in clinical practices. Instead, clinicians often verify and correct auto-localization results and prefer a system that has self-improvement ability using these corrected shapes as new training shapes.Therefore, lung localization in chest X-ray becomes an ideal use case to test the effectiveness of our online dictionary method.

Our lung localization system starts from a set of auto-detected landmarks around the lung (e.g., the bottom-left lung tip), based on which lung shapes are

inferred using shape priors. Note that various factors, e.g., imaging artifacts, lung diseases, etc., might induce missing/wrong landmark detection, which should be corrected by shape prior models. Although the overall system performance depends on multiple components, including initial landmark detection, shape prior modeling and the following deformable segmentation, our comparison focuses on the shape prior modeling part, i.e., other components remain the same in comparsions. Our experimental dataset includes 367 X-ray images from different patients. 32 of them are used as training data to construct the initial data matrix/dictionary D in Eq. 1. Note that simply stacking more training shapes into D can also improve the capability of shape representation. However, it dramatically reduces the computational efficiency, which highly depends on the scale of D when solving Eq. 1 [8].

Three shape prior methods are compared, 1) the PCA based prior as used in Active Shape Model [3], 2) SSC [11], and 3) our method. Fig. 1 shows an example of using these methods to infer shapes from auto-detected landmarks. This case is challenging due to the misplaced medical instrument, which causes erroneous detections (marked by a red box in Fig. 1). Although all

Table 1. Quantitative comparisons of the lung localization using shape priors. P, Q, DSC stand for the sensitivity, specificity, and dice similarity coefficient (%), respectively.

	P	Q	DSC
PCA	87.5 ± 5.2	96.0 ± 3.1	90.1 ± 4.0
SSC	86.7 ± 4.8	$\mathbf{96.6 \pm 2.4}$	89.4 ± 3.9
Ours	$\mathbf{94.3 \pm 4.6}$	96.2 ± 2.3	$\mathbf{94.5 \pm 3.6}$

three methods achieve reasonable accuracy, the whole shape of PCA result shifts slightly to the right (where the red arrow points in Fig. 1), because PCA is sensitive to outliers. Benefited by the sparse representation and $L1$-norm constraint, SSC and our method can both handle erroneous detections. However, since the initial shape dictionary may not be generative and representative enough, the inferred shape from SSC is not as accurate as the proposed method, which updates the dictionary on-the-fly and improves its capability of shape representations. Table 1 shows the quantitative accuracy (compared to experts' annotations) of the three methods, in terms of the sensitivity, specificity, and dice similarity coefficient (DSC). In general, our method achieves significantly better sensitivity, while slightly worse specificity than SSC. The reason is that SSC under-segments some images, which results in low sensitivity but high specificity. Our method achieves much better performance in terms of DSC, which is a more comprehensive measurements (includes both sensitivity and specificity) for localization accuracy. The experiments are performed on a PC with 2.4GHz Intel Quad CPU, 8GB memory, with Python 2.5 and C++ implementations. The whole framework is fully automatic and efficient. The shape inference step takes 0.2-0.3s, with around 0.06s as an overhead to update the dictionary online, which is negligible. In contrast, re-training the dictionary using K-SVD needs around 15-40s each time.

Real-Time Left Ventricle Tracking: Extraction of the boundary contour of a beating heart from cardiac MRI image sequences plays an important role
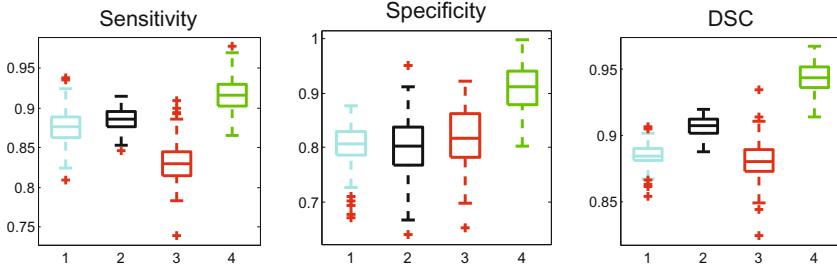
Fig. 2. Box plots for quantitative comparisons. Blue, black, red and green boxes represent results from the deformation, PCA, SSC, and the proposed method, respectively.

in cardiac disease diagnosis and treatments. MRI-guided robotic intervention is potentially important in cardiac procedures such as aortic valve repair. One of the major difficulties is the path planning of the robotic needle, which requires accurate contour segmentation of the left ventricle on a *real-time* MRI sequence. Thus, the algorithm should be robust, accurate and fast. We use a shape prior based tracking framework to solve this problem.

In our method, a collaborative trackers network is employed to provide a deformed mesh and then generate a rough contour as the initialization at each time step [12]. Next, this initialized shape model deforms based on low level image appearance. Appearance-based deformation may not be accurate since the image information can be ambiguous and noisy. Thus, the shape prior model is employed to refine the deformed contour. Based on this framework, we compare the performance of (a) deformable model based on image appearance, (b) PCA based, (c) SSC based and (d) the online dictionary based shape refinement methods. For computational efficiency consideration, the dictionary size of (b) and (c) is fixed as a small number 8. The SSC method constantly uses this initial dictionary, while the proposed method (c) updates the dictionary on-the-fly by using acquired tracking results as the mini-batch input of Algorithm 1.

Fig. 2 shows the quantitative evaluations, in terms of the sensitivity, specificity, and the dice similarity coefficient. Appearance-based deformation results produces inconsistent results when the image information is ambiguous. SSC based shape refinement may not improve the accuracy of the deformed result due to the small size of dictionary. PCA based method achieves good performance. However, it is not able to handle certain new shapes which cannot be generalized from the current PCA results. In general, the proposed method achieves the most accurate result, since it updates the dictionary on-the-fly using newly acquired information. Thus it is more generic and adaptive to new data. Online updating the dictionary takes around 0.03s, which causes very small overhead for the whole system. To track total of 189 frames, our system takes 23.7s. Retraining the dictionary using K-SVD takes around 12s each time, which is not feasible for realtime applications.

4   Conclusions

In this paper, we propose a shape modeling method to tackle the two limitations of Sparse Shape Composition (SSC). Instead of directly including all training shapes in a repository, we employ dictionary learning technologies to learn a compact and informative shape dictionary. In more details, an initial shape dictionary is learned by K-SVD using available training shapes. When new training shapes come, online dictionary learning method is used to update the dictionary on-the-fly. With the dynamic updated dictionary, SSC is gracefully scaled-up to contain shape priors from a large number of training shapes without losing the run-time efficiency. Compared to standard SSC, it achieved better shape modeling performance with a much faster speed.

References

1. Aharon, M., Elad, M., Bruckstein, A.: K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. IEEE Transactions on Signal Processing 54(11), 4311–4322 (2006)
2. Candes, E., Romberg, J., Tao, T.: Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. IEEE Transactions on Information Theory 52(2), 489–509 (2006)
3. Cootes, T., Taylor, C., Cooper, D., Graham, J.: Active shape model - their training and application. Computer Vision and Image Understanding 61, 38–59 (1995)
4. Efron, B., Hastie, T., Johnstone, I., Tibshirani, R.: Least angle regression. The Annals of Statistics 32(2), 407–499 (2004)
5. Hufnagel, H., Pennec, X., Ehrhardt, J., Handels, H., Ayache, N.: Shape Analysis Using a Point-Based Statistical Shape Model Built on Correspondence Probabilities. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part I. LNCS, vol. 4791, pp. 959–967. Springer, Heidelberg (2007)
6. Langs, G., Paragios, N., Essafi, S.: Hierarchical 3D diffusion wavelet shape priors. In: ICCV, pp. 1717–1724 (2010)
7. Mairal, J., Bach, F., Ponce, J., Sapiro, G.: Online dictionary learning for sparse coding. In: ICML, pp. 689–696 (2009)
8. Nesterov, Y., Nemirovsky, A.: Interior point polynomial methods in convex programming. Studies in Applied Mathematics 13, 1993 (1994)
9. Shi, Y., Qi, F., Xue, Z., Chen, L., Ito, K., Matsuo, H., Shen, D.: Segmenting lung fields in serial chest radiographs using both population-based and patient-specific shape statistics. IEEE Transactions on Medical Imaging 27(4), 481–494 (2008)
10. Yan, P., Kruecker, J.: Incremental Shape Statistics Learning for Prostate Tracking in TRUS. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part II. LNCS, vol. 6362, pp. 42–49. Springer, Heidelberg (2010)
11. Zhang, S., Zhan, Y., Dewan, M., Huang, J., Metaxas, D., Zhou, X.: Towards robust and effective shape modeling: Sparse shape composition. Medical Image Analysis 16(1), 265–277 (2012)
12. Zhou, Y., Yeniaras, E., Tsiamyrtzis, P., Tsekos, N., Pavlidis, I.: Collaborative Tracking for MRI-Guided Robotic Intervention on the Beating Heart. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part III. LNCS, vol. 6363, pp. 351–358. Springer, Heidelberg (2010)

Detection of Substantia Nigra Echogenicities in 3D Transcranial Ultrasound for Early Diagnosis of Parkinson Disease

Olivier Pauly[1,2], Seyed-Ahmad Ahmadi[2], Annika Plate[3],
Kai Boetzel[3], and Nassir Navab[2]

[1] Institute of Biomathematics and Biometry, Helmholtz Zentrum München, Germany
[2] Computer Aided Medical Procedures, Technische Universität München, Germany
[3] Department of Neurology, Ludwig-Maximilians-University of Munich, Germany
{pauly,ahmadi,navab}@cs.tum.edu

Abstract. Parkinson's disease (PD) is a neurodegenerative movement disorder caused by decay of dopaminergic cells in the substantia nigra (SN), which are basal ganglia residing within the midbrain area. In the past two decades, transcranial B-mode sonography (TCUS) has emerged as a viable tool in differential diagnosis of PD and recently has been shown to have promising potential as a screening technique for early detection of PD, even before onset of motor symptoms. In TCUS imaging, the degeneration of SN cells becomes visible as bright and hyper-echogenic speckle patches (SNE) in the midbrain. Recent research proposes the usage of 3D ultrasound imaging in order to make the application of the TCUS technique easier and more objective. In this work, for the first time, we propose an automatic 3D SNE detection approach based on random forests, with a novel formulation of SNE probability that relies on visual context and anatomical priors. On a 3D-TCUS dataset of 11 PD patients and 11 healthy controls, we demonstrate that our SNE detection approach yields promising results with a sensitivity and specificity of around **83%**.

1 Introduction and Medical Motivation

Parkinson's Disease (PD) is a neuro-degenerative movement disorder which has been the matter of increasing research in the medical and scientific community for the past decades. The primary symptoms of PD affect the motoric system, such as rigidity, shaking or slowness, but PD may also evoke non-motor symptoms such as dementia in later stages of the disease. The root cause of PD is the death of dopaminergic substantia nigra (SN) cells, which are located in the midbrain area. Although it is not known whether it is the cause or an effect of SN cell death, the progress of the disease is accompanied by a build-up of ferrite deposits within the SN. Over the past two decades, several studies have shown that these physiological changes can be visualized using transcranial ultrasound (TCUS), making this imaging technique a viable tool in differential diagnosis of PD [12]. Additionally, it has been shown recently that TCUS can be used as an early indicator of PD [2]. This result is particularly relevant, since

it increases the hope that TCUS can be used as a cheap, quick and non-invasive early-detection and screening tool for large populations. The changes in SN are visible in TCUS in form of hyper-echogenicities, i.e. small bright speckle patches, within the midbrain. If performed by an expert sonographer with substantial experience in this technique, sensitivity and specificity of this technique can be as high as 90% [7]). However, the challenging nature of TCUS images causes high intra- and inter-rater variability and makes it difficult for less experienced groups to reach the diagnostic reliability of expert groups in this field [11]. Recently, the usage of three-dimensional (3D-) TCUS started being investigated for PD diagnosis , since it can make this promising PD screening technique easier, more objective, and more significant due to the volumetric analysis of substantia nigra echogenicities (SNE). In this paper, we introduce a fully automatic approach for the detection of SNE voxels within the midbrain, once the latter has been localized. There is little related work in literature concerning the automatic analysis of SNE, but similar to our work, all approaches we are aware of perform a midbrain ROI segmentation first and a SNE detection within the midbrain subsequently. Kier et al. [9] and Chen et al. [4] respectively perform SN pixel detection using morphological operators or image-feature-based SVM classification, both within a manually segmented midbrain in 2D. Engels et al. [6] use a hierarchical finite-element model and active contours to simultaneously segment the midbrain and SNEs in 2D. Despite early work on segmentation of midbrain area in 3D ultrasound [1], to our knowledge, there is no previous work on (semi-) automatic SNE analysis in 3D-TCUS. The main contributions of this paper are therefore to 1) propose a novel and *volumetric* SNE detection method based on random-forest, 2) formulate a detection paradigm mimicking human experts by using probabilistic modeling of visual and spatial SNE features and 3) show the reliability of our SNE detection approach on a set of 3D-TCUS volumes from 22 subjects.

2   Materials and Methods

As illustrated by Fig. 1, an experimented observer can detect PD-related hyper-echogenicities in the left and right SN using 3D-TCUS. Unfortunately, TCUS cannot visualize the SN regions themselves, but only the high-contrast SNE speckles located randomly within the area of SN. Thus, relying on prior knowledge of the midbrain anatomy and the known rough location of the SN within the midbrain, an experimented observer has to decide whether an echogenicity belongs to the SN or not based on location and intensity of speckle patches. This makes the detection of Parkinson-related SNEs quite challenging. In the present work, we aim at providing a reliable detection of PD-related SNEs in 3D by analoguously integrating two types of information: (i) visual context and (ii) spatial location within the midbrain.

Problem Statement: Let us consider an intensity function denoted by $\mathbf{I}$: $\Omega \to \mathbb{R}$, where $\Omega \subset \mathbb{R}^3$ is the image domain representing the 3D ultrasound
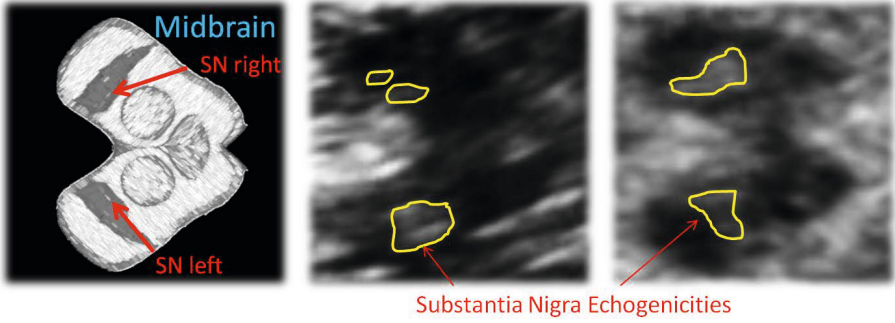
Fig. 1. Goal of our approach: On the left, the anatomy of the midbrain is detailed, showing the Substantia Nigra regions located at the front of both hemispheres. The other images show examples of typical SNE speckle patterns (in yellow) in 3D-TCUS transversal slices.

data. We further assume that we are given a segmentation of the midbrain $\mathcal{M} \subset \Omega$, either from a manual expert segmentation or alternatively from the output of a ROI detection algorithm [1]. In this paper, we propose to formulate the detection problem as a classification task in which each voxel $\mathbf{x} \in \mathcal{M}$ needs to be associated to a label $\mathbf{c} \in \{0, 1\}$, where 0 denotes the background and 1 the Substantia Nigra Echogenicities (SNE) class. In fact, $\mathbf{c}$ is the realization of 2 random variables $(\mathcal{E}, \mathcal{S})$ where $\mathcal{E}$ represents the observation of an echogenicity and $\mathcal{S}$ of the Substantia Nigra (SN), i.e. $\mathbf{c} = 1$ if and only if $\mathcal{E} = 1$ **and** $\mathcal{S} = 1$. Therefore, we aim at learning $P(\mathcal{E}, \mathcal{S}|\mathbf{x}, \mathbf{I})$, which represents the joint probability of observing an echogenicity $\mathcal{E}$ belonging to the SN $\mathcal{S}$ given the location $\mathbf{x}$ and the intensity function $\mathbf{I}$. It is important to note that (1) it is not the SN itself which causes hyper-echogenicities but only potential acoustic micro-scatterers residing within it and (2) echogenicities can happen in the whole skull in TCUS due to tissue boundaries and micro-scatterers present in the entire brain tissue. Hence, we can assume the independence of the random variables $\mathcal{E}$ and $\mathcal{S}$, and decompose this joint probability as follows:

$$P(\mathcal{E}, \mathcal{S}|\mathbf{x}, \mathbf{I}) = P(\mathcal{E}|\mathbf{x}, \mathbf{I})P(\mathcal{S}|\mathbf{x}) \tag{1}$$

The first term $P(\mathcal{E}|\mathbf{x}, \mathbf{I})$ is a data term, encoding the probability of observing an echogenicity given some visual information at location $\mathbf{x}$, and the second term $P(\mathcal{S}|\mathbf{x})$ is an anatomical prior not depending on $\mathbf{I}$, i.e. the ultrasound data. As learning these probability distributions is challenging due to the dimensionality of the problem, we propose to use two discriminative models based on random forests. Following a "divide" and "conquer" strategy, random forests [3] provide efficient piecewise approximations of any distribution in high-dimensional spaces by: (1) partitioning the space using simples decisions, and (2) estimating the posterior in each "cell" of this space. As shown in [5,10] , random forests have been successfully applied to the task of multiple organ localization in CT and MR scans. Geremia et al. in [8] demonstrated state-of-the-art results for the

segmentation of multiple-sclerosis lesions based on multi-channel MRI data. In addition to a forest using visual context, we propose to learn a novel spatial prior based on two hemisphere-specific coordinate systems. In the following, we describe how to use random forests for learning: (1) the data term $P(\mathcal{E}|\mathbf{x}, \mathbf{I})$ and (2), the prior $P(\mathcal{S}|\mathbf{x})$.

Learning the Data Term $P(\mathcal{E}|\mathbf{x}, \mathbf{I})$: In TCUS, echogenicities are character-ized by higher intensities and higher contrast. Therefore, we propose to describe the visual context of a voxel at location $\mathbf{x}$ by extracting a set of simple features that encode the mean intensities in cuboidal regions of different sizes in the neighborhood of $\mathbf{x}$ similarly as in [8]. Let us denote by $\mathcal{X}$ the space spanned by these simple features, and $\mathbf{X}$ the feature representation associated to a voxel at location $\mathbf{x}$. We consider a training set $(\mathbf{X}_n, \mathcal{E}_n)_{n=1}^{N}$, where each feature vector $\mathbf{X}_n$ is associated to a label $\mathcal{E}_n$ which is equal to 1 if there is an echogenicity at location $\mathbf{x}_n$ and 0 if not. Consisting of an ensemble of independent trees, a random forest permits to efficiently partition this high-dimensional space $\mathcal{X}$. Each tree can be seen as a directed acyclic graph where each node consists in a decision function $f_{\mathbf{v}, \tau}$ defined as $f_{\mathbf{v}, \tau}(\mathbf{X}) = (\mathbf{X} \cdot \mathbf{v} \geq \tau)$. $\mathbf{v}$ is a vector of dimensionality $\dim(\mathcal{X})$ having only 1 non-zero value, and $\tau \in \mathbb{R}$ a threshold. According to the result of this decision function, incoming data are pushed towards the left or right child of the current node. Note that the role of $\mathbf{v}$ is to select a feature dimension where to perform the decision, yielding thus axis-aligned splits in $\mathcal{X}$. Let us denote by Δ the set of feature points from $\mathcal{X}$ reaching the current node, and Δ_l, Δ_r the subsets respectively sent to the left and right child nodes. At each node, the choice of $\mathbf{v}$ and τ is optimized following a greedy optimization strategy. A set Γ of functions are randomly drawn and the best candidate $(\mathbf{v}^*, \tau^*)$ is selected by maximizing information gain:

$$(\mathbf{v}^*, \tau^*) = \underset{(\mathbf{v}, \tau) \in \Gamma}{\operatorname{argmax}} (\mathbf{H}(\Delta) - w_l \mathbf{H}(\Delta_l) - w_r \mathbf{H}(\Delta_r)) \qquad (2)$$

where $w_l = |\Delta_l|/|\Delta|$ and $w_r = |\Delta_r|/|\Delta|$. $\mathbf{H}$ corresponds to the classical Shan-non's entropy $\mathbf{H} = -\sum_{e \in \{0,1\}} P(\mathcal{E} = e|\mathbf{x}, \mathbf{I}) \log (P(\mathcal{E} = e|\mathbf{x}, \mathbf{I}))$, the posterior distribution being estimated from the set of points in the current node as:

$$P(\mathcal{E} = e|\mathbf{x}, \mathbf{I}) = \frac{|\{\mathbf{X}_n \in \Delta, \ \mathcal{E}_n = e\}|}{|\{\mathbf{X}_n \in \Delta\}|} \qquad (3)$$

By optimizing this energy function, the tree aims at minimizing the uncertainty on the random variable $\mathcal{E}$, encouraging thereby the creation of leaves containing either mostly echogenicities, or mostly background. Nodes are grown until a maximal tree depth has been reached, or when the number of feature points falls below a given threshold. Finally, in each leaf, the posterior distribution $P(\mathcal{E}|\mathbf{x}, \mathbf{I})$ is computed on the set of features points reaching this leaf using Eq.3 and stored. Now, to predict the probability of observing an echogenicity at a location $\mathbf{x}$ for an unseen ultrasound volume of the midbrain, one just needs to first extract its associated feature vector $\mathbf{X}$, to push it downward the tree until it

Midbrain in TC-US Transversal plane   Left hemishere / Right hemisphere   Hemisphere center/Coordinate system
Substantia Nigra   Voxel relative coordinates
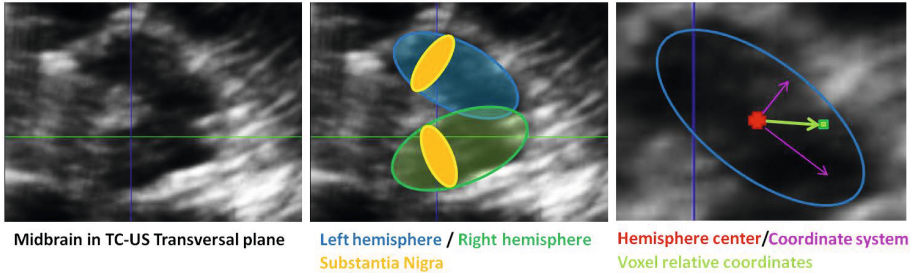
Fig. 2. Midbrain anatomy: in the transversal plane, the midbrain has a characteristic butterfly shape. The Substantia Nigra are thin structures located at the front of both hemispheres. A hemisphere-specific coordinate system is computed to express voxel spatial location accounting for inter-patient asymmetric changes of scales and orientation.

reaches a leaf, and to use the stored posterior distribution. Considering a random forest consisting of T trees, predictions can be simply computed by averaging tree posteriors: $P(\mathcal{E}|\mathbf{x}, \mathbf{I}) = \frac{1}{T} \sum_t^T P_t(\mathcal{E}|\mathbf{x}, \mathbf{I})$.

Learning the Prior $P(\mathcal{S}|\mathbf{x})$: As shown on Fig. 2, the midbrain has a characteristic butterfly shape in the transversal plane, which does not vary much along the longitudinal axis. The Substantia Nigra are thin structures located at the front of both hemispheres and do not vary much along the longitudinal axis either. Hence, we propose to express the location of each voxel using patient-specific coordinate systems that represent the left and right midbrain hemispheres in the transversal plane. By doing so, we can easily account for asymmetric changes of scales and orientation of the midbrain anatomy, which can occur in TCUS imaging. Let us denote by $\{\mathbf{x}_m\}_{m=1}^M = \mathcal{M}$, the finite set of M voxels belonging to the midbrain. First, the centers of the left and right hemispheres are computed by performing a K-means clustering on $\mathcal{M}$. Then, each voxel is associated to its nearest cluster center to create the 2 hemisphere subsets $\mathcal{H}^{\text{left}}$ and $\mathcal{H}^{\text{right}}$. Finally, principal component analysis is applied to each of these subsets to compute a hemisphere-specific transversal coordinate system, and the location of each point is expressed in the normalized coordinate systems of the hemisphere it belongs to. The in-plane location of each voxel $\mathbf{x}_m$ can then be encoded by a vector $\mathbf{x}'_m = [x'_m, y'_m, h_m]$, where x'_m and y'_m are the in-plane components in the hemisphere coordinate system, and h_m is a categorical variable encoding the left/right side. To summarize, each voxel $\mathbf{x}_m$ is associated for the training phase to a couple $(\mathbf{x}'_m, \mathcal{S}_m)$, where $\mathcal{S}_m$ is equal to 1 if $\mathbf{x}_m$ belongs to the Substantia Nigra and 0 if not. As in the previous section, we use a random forest to learn the prior $P(\mathcal{S}|\mathbf{x})$ using a training set of 3D-TCUS from different patients. During the training, each tree aims at separating the SN from the rest of the midbrain, and creates clusters in its leaves that are consistent in terms of spatial location $\mathbf{x}'_m$.
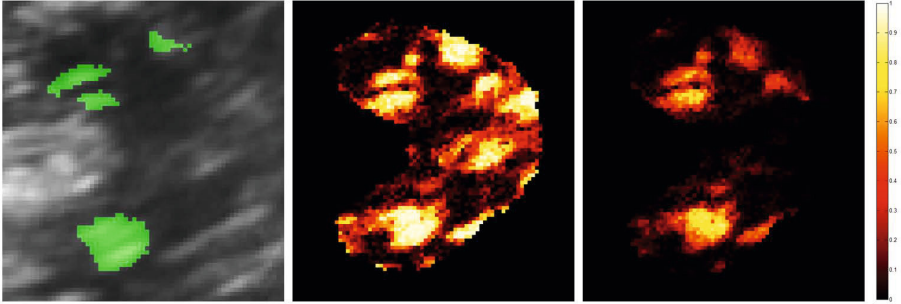
Fig. 3. The effect of our spatial prior: From left to right, (i) the manual segmentation overlaid on the US data, (ii) the predicted posterior using the data term forest and (iii) the output after combining with the forest-based spatial prior. All outputs are probabilistic and can be thresholded to provide a binary segmentation.

SNE Detection: Once the data term and the prior have been learned from a set of labelled midbrains, a new unseen patient data can be processed as follows: (1) the midbrain is segmented, (2) the hemisphere coordinate systems are determined using K-means followed by a PCA, (3) the probability $P(\mathcal{E}|\mathbf{x}, \mathbf{I})$ and the prior $P(\mathcal{S}|\mathbf{x})$ are computed for each voxel, and (4) the joint probability $P(\mathcal{E}, \mathcal{S}|\mathbf{x}, \mathbf{I})$ can be predicted using Eq. 1. Hence, we obtain for each voxel a probability of belonging to an SNE, and we can use a threshold $\mathcal{T} \in [0, 1]$ to create a binary segmentation of the ferrite deposits: $\mathbf{c} = 1$ if $P(\mathcal{E}, \mathcal{S}|\mathbf{x}, \mathbf{I}) \geq \mathcal{T}$, and $\mathbf{c} = 0$ otherwise.

3   Experiments and Results

In this section, we evaluate our SNE detection approach on the bi-lateral 3D-TCUS dataset volume of 22 subjects, consisting of 11 PD patients and 11 healthy controls. The 3D volumes were reconstructed at an isotropic resolution of 0.45mm and labelled by a blinded expert into the regions "midbrain", "SNE left" and "SNE right". For our validation, we will consider this labeling as gold standard. We conduct comparative experiments to evaluate our SNE detection approach based on 2 discriminative models (VisForest-PriorForest) against the simple forest without spatial prior (VisForest), and a forest with a spatial prior constructed using a Gaussian distribution model for each hemisphere (VisForest-GaussianPrior). The parameters of each Gaussian spatial prior are estimated by computing the sample mean and the covariance of the location of the SNE voxels in their hemisphere coordinate systems.

We perform a leave-one-patient-out cross-validation (LOO), i.e. we train all models on 21 labelled midbrains and test on the remaining one. As the outputs from our system are probabilities between 0 and 1, we perform a ROC analysis, i.e. we vary the threshold's value to compute a binary segmentation, compute the corresponding confusion matrices for each run and derive different quality

Table 1. Overall SNE Detection results on 22 patients: The proposed prior permits to achieve better detecton by improving the specificity, i.e. by better rejecting echogenicities that do not belong to the estimated SN. Moreover, using a forest-based prior provides slightly better results.

	F-measure			Specificity			Sensitivity		
	Mean	Std	Median	Mean	Std	Median	Mean	Std	Median
VisForest	0.456	0.115	0.463	0.775	0.060	0.779	**0.845**	**0.081**	**0.859**
VisForest-GaussianPrior	0.508	0.155	0.547	0.819	0.045	0.812	0.829	0.113	0.844
VisForest-PriorForest	**0.519**	**0.148**	**0.574**	**0.835**	**0.043**	**0.832**	0.828	0.099	0.829

measures: f-measure, specificity and sensitivity. While the number of trees is set to 10 for all experiments, the depth parameter is tuned by doing a discrete search (i.e. depth = 5,10,15,20) and performing a full LOO for each depth value. Best results were obtained for a depth = 15 for the VisForest, and for a depth = 10 for the PriorForest.

Overall results are presented in Tab. 1. On the left, the best f-measure are reported by using threshold values of 0.5, 0.1 and 0.2 respectively for the Vis-Forest, VisForest-GaussianPrior and VisForest-PriorForest models. By including our hemisphere-specific spatial prior, the f-measure is increased from **0.456** (Vis-Forest) to **0.518** (VisForest-PriorForest). Moreover, learning this prior distribution using a random forest provides slightly better results than with Gaussian prior achieving **0.508**. On the right, the best compromise between sensitivity and specificity are computed from the ROC analysis for all approaches. As illustrated by Fig. 3, the proposed prior permits to achieve improved specificity by better rejecting echogenicities that do not belong to the estimated SN. By varying the segmentation threshold, we also compute the area under curve which is **AUC = 0.903** for our approach, compared to a VisForest alone **AUC = 0.879** or with a simple Gaussian prior **AUC = 0.891**.

4   Discussion and Conclusion

In this paper, we presented the first approach for the automatic detection of Substantia Nigra Echogenicities in 3D-TCUS. As the interpretation of such data is very difficult and yields high inter and intra-observer variability, our aim is to provide an objective and reliable segmentation of such Parkinson-related speckle patches. Inspired by the way medical experts recognize SNE, we proposed a probabilistic formulation combining two discriminative models: (1) a "visual" random forest specialized on the detection of echogenicities and (2) a "spatial" random forest modeling a location prior within the midbrain. Therefore, voxel locations are parametrized within hemisphere-specific coordinate systems in order to account for asymmetric changes of orientation and scale in the midbrain anatomy. In experimentations conducted on 22 patients data, we could assess the reliability of our approach that achieves a sensitivity and specificity of around **83%**. From the segmentation output of our system, we can quantify automatically the

amount of hyper-echogenicities in each hemisphere, with the motivation of using this information within a computer aided diagnosis system for Parkinson disease based on 3D-TCUS in the near future.

References

1. Ahmadi, S.-A., Baust, M., Karamalis, A., Plate, A., Boetzel, K., Klein, T., Navab, N.: Midbrain Segmentation in Transcranial 3D Ultrasound for Parkinson Diagnosis. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part III. LNCS, vol. 6893, pp. 362–369. Springer, Heidelberg (2011)
2. Berg, D., Seppi, K., Behnke, S., Liepelt, I., Schweitzer, K., Stockner, H., Wollenweber, F., Gaenslen, A., Mahlknecht, P., Spiegel, J., Godau, J., Huber, H., Srulijes, K., Kiechl, S., Bentele, M., Gasperi, A., Schubert, T., Hiry, T., Probst, M., Schneider, V., Klenk, J., Sawires, M., Willeit, J., Maetzler, W., Fassbender, K., Gasser, T., Poewe, W.: Enlarged substantia nigra hyperechogenicity and risk for Parkinson disease: a 37-month 3-center study of 1847 older persons. Arch. Neurol. 68, 932–937 (2011)
3. Breiman, L.: Random forests. Machine Learning 45(1), 5–32 (2001)
4. Chen, L., Seidel, G., Mertins, A.: Multiple Feature Extraction for Early Parkinson Risk Assessment Based on Transcranial Sonography Image. In: IEEE Int. Conf. on Image Processing (2010)
5. Criminisi, A., Shotton, J., Konukoglu, E.: Decision forests: A unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning. In: Foundations and Trends in Computer Graphics and Vision (2012)
6. Engel, K., Toennis, K.D.: Segmentation of the Midbrain in Transcranial Sonographies using a Two–Component Deformable Model. Annals of the British Machine Vision Association and Society for Pattern Recognition (BMVA) 4, 1–12 (2009)
7. Gaenslen, A., Unmuth, B., Godau, J., Liepelt, I., Di Santo, A., Schweitzer, K.J., Gasser, T., Machulla, H.J., Reimold, M., Marek, K., Berg, D.: The specificity and sensitivity of transcranial ultrasound in the differential diagnosis of Parkinson's disease: a prospective blinded study. Lancet Neurol. 7, 417–424 (2008)
8. Geremia, E., Menze, B.H., Clatz, O., Konukoglu, E., Criminisi, A., Ayache, N.: Spatial Decision Forests for MS Lesion Segmentation in Multi-Channel MR Images. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part I. LNCS, vol. 6361, pp. 111–118. Springer, Heidelberg (2010)
9. 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. Int. J. of Computer Assisted Radiology and Surgery 2(S1), S83–S85 (2007)
10. Pauly, O., Glocker, B., Criminisi, A., Mateus, D., Möller, A.M., Nekolla, S., Navab, N.: Fast Multiple Organ Detection and Localization in Whole-Body MR Dixon Sequences. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part III. LNCS, vol. 6893, pp. 239–247. Springer, Heidelberg (2011)
11. 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. BioMed Central Neurology 8, 42 (2008)
12. 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. Arch. Neurol. 64(11), 1635–1640 (2008)

Prostate Segmentation by Sparse Representation Based Classification

Yaozong Gao[1,2], Shu Liao[2], and Dinggang Shen[2]

[1] Department of Computer Science, University of North Carolina at Chapel Hill
yzgao@cs.unc.edu
[2] Department of Radiology and BRIC, University of North Carolina at Chapel Hill
liaoshu.cse@gmail.com, dgshen@med.unc.edu

Abstract. Accurate segmentation of prostate in CT images is important in image-guided radiotherapy. However, it is difficult to localize the prostate in CT images due to low image contrast, unpredicted motion and large appearance variations across different treatment days. To address these issues, we propose a sparse representation based classification method to accurately segment the prostate. The main contributions of this paper are: (1) A discriminant dictionary learning technique is proposed to overcome the limitation of the traditional Sparse Representation based Classifier (SRC). (2) Context features are incorporated into SRC to refine the prostate boundary in an iterative scheme. (3) A residue-based linear regression model is trained to increase the classification performance of SRC and extend it from hard classification to soft classification. To segment the prostate, the new treatment image is first rigidly aligned to the planning image space based on the pelvic bones. Then two sets of location-adaptive SRCs along two coordinate directions are applied on the aligned treatment image to produce a probability map, based on which all previously segmented images of the same patient are rigidly aligned onto the new treatment image and majority voting strategy is further adopted to finally segment the prostate in the new treatment image. The proposed method has been evaluated on a CT dataset consisting of 15 patients and 230 CT images. Promising results have been achieved.

1   Introduction

Prostate cancer is the second-leading cancer for American men. Currently one of the major treatment methods is external beam radiation therapy, which basically has two stages, namely *the planning stage* and *the treatment stage*. In the planning stage, *a planning image* is scanned from the patient and a dose plan is designed. During the treatment stage, a CT image is acquired at each treatment day for the same patient, which could be repeated for up to 40 times, each with a CT image acquired. The prostate in each treatment image needs to be accurately localized so that the dose plan made in the planning image can be adjusted to the current treatment image. Therefore, the success of external beam radiation therapy highly depends on the accurate localization of the prostate.
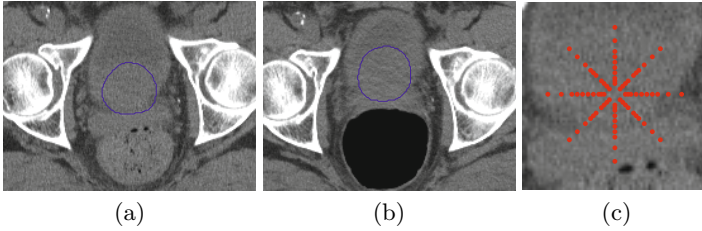
Fig. 1. (a) and (b) are two axial slices from different treatment images of the same patient. Blue contours are the prostate contours manually delineated by experts. (c) is an illustration of context locations of the center pixel by red points.

However, there are three main challenges to accurately segment the prostate. First, prostate boundary is of extremely low contrast with its surrounding tissues in the CT images as shown in Fig. 1. Second, prostate motion is unpredictable due to the uncertain existence of bowel gas in different treatment days. Third, the bowel gas can significantly alter the image appearance and makes it inconsistent across different treatment days as illustrated in Fig. 1(a) and 1(b). In order to address these challenges, many novel methods have been proposed these years. The first category of methods is deformable model [1, 2]. The second category is registration-based method [3, 4]. Recently, Li et al. [5] incorporated context features into prostate segmentation and achieved promising results.

On the other hand, Sparse Representation based Classifier (SRC) [6] has achieved the state-of-the-art results in face recognition. It represents a testing sample as a sparse linear combination with respect to an over-complete dictionary, which consists of training samples from all classes. Representation residue with respect to each class is used to determine the class label of a testing sample. However, the good performance of the traditional SRC depends on the assumption that the training samples in each class are distinct from those in other classes. In practice, especially in pixel-wise classification, different classes may include very similar samples. To overcome this limitation, we propose a discriminant dictionary learning technique to enhance the dissimilarity between classes. Moreover, context features are incorporated into SRC to refine the prostate boundary in an iterative scheme. Finally, a linear regression model is further trained to predict the class probability based on the representation residues. The proposed method has been evaluated on 230 CT images from 15 patients. The experimental results show that our method can achieve promising results and outperform other state-of-the-art prostate segmentation methods.

2 Methodology

2.1 Sparse Representation Based Classifier (SRC)

Given a dictionary $D \in \mathbb{R}^{n \times N}$ and a sample $\mathbf{y} \in \mathbb{R}^n$, sparse representation aims to find a sparse linear combination of dictionary elements in D for best

representing $\mathbf{y}$. Mathematically, the problem can be formulated as the following minimization:

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0, \text{ subject to } \|\mathbf{y} - D\mathbf{x}\|_2 \leq \varepsilon, \qquad (1)$$

where $\mathbf{x}$ contains the sparse linear coefficients and is usually called as sparse code in the literature, and ε is the maximum allowable representation error. Although solving (1) is a NP-hard problem, the solution can be well approximated by many pursuit algorithms such as Basis Pursuit (BP) and Orthogonal Matching Pursuit (OMP) [7]. In consideration of both efficiency and performance, we use OMP to solve the sparse coding problem in this paper.

In the traditional SRC, the dictionary D is formed as a collection of training samples from all classes:

$$D = [D_1, D_2, \cdots, D_i, \cdots, D_K] = [\mathbf{d}_{1,1}, \mathbf{d}_{1,2}, \cdots, \mathbf{d}_{i,j}, \cdots, \mathbf{d}_{K,N_K}], \qquad (2)$$

where D_i is a sub-dictionary of class i that only contains training samples from class i, $\mathbf{d}_{i,j}$ is the j-th training sample of class i, K is the number of classes and N_K is the number of training samples in class K. To classify a new sample $\mathbf{y}$, its sparse code $\mathbf{x}$ is first computed according to (1). Then the representation residual vector $\mathbf{r}_i$ with respect to class i is computed as:

$$\mathbf{r}_i = \mathbf{y} - D_i\mathbf{x}_i, \qquad (3)$$

where $\mathbf{x}_i$ carries entries of $\mathbf{x}$ indexed by an index set, which contains indices of columns in D belonging to D_i. Finally, the new sample $\mathbf{y}$ is classified to the class with the minimum $\|\mathbf{r}_i\|_2$.

2.2 Discriminant Sub-dictionary Learning

The traditional SRC method works well when there are no similar elements between sub-dictionaries. However, in many cases this assumption doesn't hold. In order to overcome this limitation, we need to build discriminant sub-dictionaries whose elements are distinct from those in other sub-dictionaries. In this paper, we combine feature selection with dictionary learning technique to learn a discriminant sub-dictionary for each class. Here only the case of two classes is illustrated since a voxel is either classified to object (prostate) or background in the case of prostate segmentation. But this idea can be readily extended to multi-class cases when combined with multi-class feature selection techniques.

Given both background and object training samples, we want to select the topmost discriminant features that can enlarge the dissimilarities between training samples of two classes. Feature ranking based on Fisher's Separation Criteria (FSC) [10] is adopted in this paper to select the most discriminant features while eliminate features that are similar in both classes.

After feature selection, background and object samples can be directly used to form two sub-dictionaries. However, in practice the number of samples is usually large. In consideration of both sparse coding efficiency and dictionary

storage, we need to use dictionary learning technique to learn a compact representation of training samples. In this paper, we adopt K-means as a way to learn sub-dictionaries. Compared with many reconstruction-oriented dictionary learning methods such as K-SVD [8] which don't consider discriminability during dictionary optimization, K-means can identify the individual clusters of different classes and thus can better preserve the dissimilarity between background and object class during dictionary learning. Therefore, it is more suitable when combined with SRC in classification.

2.3   Boundary Refinement by Context Features

In order to accurately localize the prostate boundary, it is necessary to draw more background and object training samples near the prostate boundary. However, these background and object samples are quite similar even after feature selection basically for two reasons: First, these samples are spatially close and sometimes next to each other. Second, prostate boundary in CT images is of extremely low contrast. To the best of our knowledge, no effective features which can accurately localize the prostate boundary have been identified. Therefore, even after performing discriminant sub-dictionary learning strategy, the SRC method can still produce many classification errors along the prostate boundary, which results in a zigzag boundary, as shown in Fig. 2(a).

Motivated by [5], we incorporate context features into SRC and propose an iterative SRC classification scheme. For each voxel, its features include not only local features but also context features taken at context locations as illustrated in Fig. 1(c). Previous classification results at context locations are used as context probability features which help guide the boundary refinement in the next classification iteration. Assume no prior information is available, we start with an uniform probability map. These context features don't help in the first iteration since they are filtered out by feature selection. However, in the later classification iterations, when the probability map becomes clearer and clearer, more context features will be selected to guide the classification refinement. Usually after several iterations, the prostate boundary becomes more refined as shown in Fig. 2(c).

| (a) | (b) | (c) |
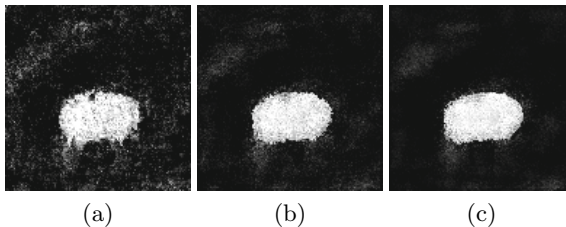
Fig. 2. The first, second and third column represents the results of the first, second and third classification iteration, respectively

2.4   Prediction by Residue-Based Linear Regression Model

The traditional SRC compares residual norms of different classes to determine the class label of a testing sample. In such case, residues of different features are equally treated. Usually a voxel is represented by the combination of different types of features, the discriminabilities of individual features are different and their contributions to classification are also different. Therefore, equally weighting them in determining the class label limits the classification performance. Besides, the traditional SRC is a hard classification method, which only assigns class label to a new sample. In contrast, soft classification provides more quantitative informaton, especially in the decision margin where the class membership is unclear.

Motivated by these observations, a residue-based linear regression model is trained to learn the contributions of different features in class probability prediction. For each training sample, its background residual vector $\mathbf{r}_0$ and object residual vector $\mathbf{r}_1$ are computed and stacked into a single vector $\mathbf{r} = [\mathbf{r}_0^{\mathrm{T}} \mathbf{r}_1^{\mathrm{T}}]^{\mathrm{T}}$, which is used together with its class label $l \in \{-1, 1\}$ to train a linear regression model $\mathbf{m} \in \mathbb{R}^{2t}$, where t is the number of selected features for each sample. For a new testing sample $\mathbf{y}_{new}$, its object (prostate) class probability is computed as:

$$p = g(\frac{\mathbf{m}^{\mathrm{T}} \mathbf{r}_{new} + 1}{2}), \tag{4}$$

where $\mathbf{r}_{new}$ is the stacked residual vector of $\mathbf{y}_{new}$ and $g(.)$ is defined as a piecewise function that maps any value outside $[0, 1]$ to its nearest boundary value in order to keep the predicted probability between $[0, 1]$.

2.5   Iterative Prostate Segmentation by SRC

We believe patches repeat not only spatially but also longitudinally. Therefore, in prostate segmentation, patches in the new treatment image likely have appeared in the previous treatment images or the planning image. If we build two discriminant patch-based sub-dictionaries for prostate and background using previous images, for a new patch in the new treatment image, it tends to draw more supports from the respective sub-dictionary in the sparse representation. Based on the representation residues corresponding to each class, we can estimate class probability of the voxel associated with this patch.

Our segmentation method consists of two stages, namely *training stage* and *classification stage*. **In the training stage**, two sets of location-adaptive [5] SRCs along two coordinate directions are learned using previous images of the same patient, which take the variability of different prostate regions into account. For each location-adaptive SRC, it only draws training samples from slices that it is responsible for. Then, based on the training samples, discriminant features are selected, two discriminant sub-dictionaries are constructed, and a residue-based linear regression model is finally learned. All these three steps are used together to classify all training slices and the class probabilities after classification are used to update the corresponding context probability features of the training

samples. After the training samples are updated, we can learn a new SRC for the next classification iteration. The process is repeated until a specified number of iterations have been reached.

In the classification stage, the middle slice along each of two coordinate directions needs to be manually specified by users in order to shift the learned SRCs to the new treatment image space for classification (Note that the automatic middle slice identification method will be developed in our future work). The classification results along two coordinate directions are fused to form a final probability map. After classification is done, all previously segmented prostate images of the same patient are rigidly aligned to the probability map of the new treatment image and then majority voting strategy is adopted to segment the prostate finally.

3   Experimental Results

Our dataset consists of 15 patients, each with more than 11 CT images, with total 230 CT images. The resolution of each CT image is 1mm×1mm×3mm. The expert manual segmentation results are available for each image to serve as the ground truth. We use the first 3 images including the planning image to initialize our method. As more treatment images are collected, only the latest 5 images are used as training images, which account for the tissue appearance change under radiation treatment. Two sets of location-adaptive SRCs are placed along anterior-posterior (y) direction and superior-inferior (z) direction, respectively, because slices along these two directions contain richer context information (e.g., pelvic bones) than slices along lateral (x) direction.

Before any operation is applied in the training stage, all previous treatment images are rigidly aligned to the planning image based on the pelvic bone structures in order to remove the irrelevant whole-body motion. The same preprocessing is also applied to the new treatment image before classification.

For each voxel, its features include both local appearance features and context features. Context features have two types of features, namely context probability features and context appearance features. Context probability features have been introduced in the previous section. They are used to refine classification results and updated in each iteration. In the experiment we only use 3 classification iterations. Context appearance features are the same kinds of features as local appearance features, but taken at context locations. 9 dimensional Histogram of Oriented Gradient (HOG) [9] and 23 Haar features computed in a 21mm×21mm local window are used as appearance features in this paper.

The box-and-whisker plot of the DICE measures and centroid distances along three coordinate directions of our method are shown in Fig. 3. Fig. 4 visually compares the segmentation results using residual norm comparison and residue-based linear regression. It can be seen that in the beginning and ending slices where the prostate is relatively small and difficult to accurately localize, the proposed linear regression model performs better than the traditional residual norm comparison. Four existing state-of-the-art prostate segmentation methods
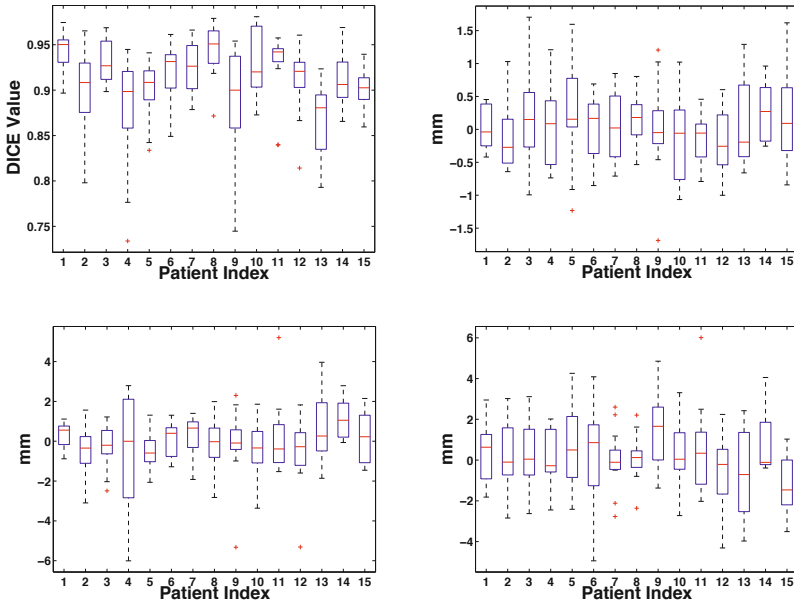
Fig. 3. Left-top figure shows the DICE measures of our method. Right-top, left-bottom and right-bottom figures are centroid distances in lateral (x), anterior-posterior (y) and superior-inferior (z) directions, respectively.
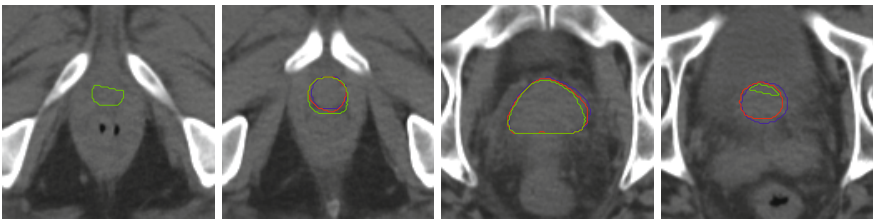
Fig. 4. Comparison of the segmentation results between linear regression and residual norm comparison. Blue contours are the prostate boundaries manually delineated by experts. Red and green contours are the segmentation results of the proposed method with linear regression and residual norm comparison, respectively. This indicates that our proposed method with linear regression achieves better results, especially in the beginning and ending slices of the prostate.

[1–3, 5] are compared with our method. The mean and standard deviation of DICE measures of our method is 0.912 ± 0.044 based on Fig. 3, which is better than 0.820 ± 0.060 in [3], 0.893 ± 0.050 in [2] and 0.908 in [5]. The median DICE measure of our method is 0.918, which is also better than 0.840 in [3] and 0.906 in [2]. The median probability of detection and false alarm of our method are 0.913 and 0.072, respectively, which are better than 0.840 and 0.130 reported in [1], and 0.900 and 0.100 reported in [5]. Besides, we also compared the centroid

distances. The mean centroid distances along lateral (x), anterior-posterior (y) and superior-inferior (z) direction of our method in Fig. 3 are 0.06 mm, -0.07 mm and 0.19 mm, respectively, which are much better than the respective centroid distances of -0.26 mm, 0.35 mm and 0.22 mm reported in [3], and comparable to the result of 0.18 mm, -0.02 mm and 0.57 mm in [5].

4 Conclusion

We have proposed a sparse representation based classification method for segmentation of prostate in CT images. Feature selection is combined with dictionary learning technique to learn two discriminant sub-dictionaries which overcome the limitation of the traditional SRC. Context features are further incorporated into SRC to refine the classification results (especially the prostate boundary) in an iterative scheme. A residue-based linear regression model is finally learned to increase the classification performance and extend the traditional SRC from hard classification to soft classification. Experimental results show that our proposed method can achieve more accurate prostate segmentation results than other state-of-the-art segmentation methods under comparison.

References

1. 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)
2. 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)
3. 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. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3749, pp. 442–450. Springer, Heidelberg (2005)
4. Liao, S., Shen, D.: A Learning Based Hierarchical Framework for Automatic Prostate Localization in CT Images. In: Madabhushi, A., Dowling, J., Huisman, H., Barratt, D. (eds.) Prostate Cancer Imaging 2011. LNCS, vol. 6963, pp. 1–9. Springer, Heidelberg (2011)
5. Li, W., Liao, S., Feng, Q., Chen, W., Shen, D.: Learning Image Context for Segmentation of Prostate in CT-Guided Radiotherapy. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part III. LNCS, vol. 6893, pp. 570–578. Springer, Heidelberg (2011)
6. Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., Ma, Y.: Robust Face Recognition via Sparse Representation. PAMI 31(2), 210–227 (2009)
7. Tropp, J.A.: Greed is good: algorithmic results for sparse approximation. IEEE Transactions on Information Theory 50(10), 2231–2242 (2004)
8. Aharon, M., Elad, M., Bruckstein, A.: K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation. IEEE Transactions on Signal Processing 54(11), 4311–4322 (2006)
9. Dalal, N., Triggs, B.: Histograms of Oriented Gradients for Human Detection. In: Computer Vision and Pattern Recognition, vol. 1, pp. 886–893 (2005)
10. Guyon, I., Elisseeff, A.: An introduction to variable and feature selection. J. Mach. Learn. Res. 3, 1157–1182 (2003)

Co-segmentation of Functional and Anatomical Images

Ulas Bagci[1], Jayaram K. Udupa[2], Jianhua Yao[1], and Daniel J. Mollura[1]

[1] Center for Infectious Diseases Imaging, Department of Radiology and Imaging
Sciences, National Institutes of Health (NIH), Bethesda, USA
[2] Department of Radiology, University of Pennsylvania, Philadelphia, PA, USA

Abstract. This paper presents a novel method for segmenting functional and anatomical structures simultaneously. The proposed method unifies domains of anatomical and functional images (PET-CT), represents them in a product lattice, and performs simultaneous delineation of regions based on a random walk image segmentation. In addition, we propose a simple yet efficient object/background seed localization method, where background and foreground object cues are automatically obtained from PET images and propagated onto the corresponding anatomical images (CT). In our experiments, abnormal anatomies on PET-CT images from human subjects are segmented synergistically by the proposed fully automatic co-segmentation method with high precision (mean DSC of 91.44%) in seconds (avg. 40 seconds).

Keywords: Joint Segmentation, PET-CT, Random Walk, Object Detection.

1 Introduction

Hybrid imaging modalities such as PET-CT and MRI-PET are in vogue since they can achieve higher sensitivity and specificity than the component modalities alone [1]. As a functional measurement, standardized uptake value (SUV) is often used in Positron Emission Tomography (PET) imaging (See Figs. 1a-e). However, SUVs alone are not enough in a PET-CT acquisition to diagnose, characterize, and stage the disease, since anatomic boundaries of the corresponding structure on Computed Tomography (CT) are also needed for this calculation. As a result, diagnostic sensitivity and specificity achieved in hybrid imaging modalities (i.e., PET-CT) are higher than either modality alone. All these processes require precise segmentations of both PET and CT images. There are several reasons why region delineations need to be performed meticulously and accurately in both CT and PET images. (1) errors in segmentation can distort the SUV calculations by altering the region's margins, (2) manual or semi-automated mechanisms of segmentation can be inefficient and suffer from unacceptable inter-observer variance, and (3) using CT for segmenting lesions that are quantitatively measured on PET can overlook and exclude other important quantitative data such as texture features in PET images. Therefore, we

aim to produce an automated, efficient, and reproducible segmentation method that simultaneously unifies anatomic and functional information. The proposed method can be especially useful in quantifying lesions characterized by fuzzy boundaries and low contrast from surrounding normal structures.

Related Works: Except for a few studies [2, 3], radiotracer uptake regions are usually delineated manually in clinical routines. Fixed and adaptive thresholding and region based segmentation methods (i.e., fuzzy c-means, region growing, and watershed) are also used to determine boundaries [4]. However, segmentations of PET images in these studies are formulated without incorporating corresponding anatomical information (i.e., CT), and the accuracy, robustness, and reproducibility of these methods are suspect in more difficult cases.

A joint PET-CT image segmentation method was proposed recently in [2], where a Markov Random Field (MRF) algorithm was formulated on a graph. The method requires user interaction, and it was used only in images from head-neck with large tumors. Its performance in small uptake regions was not assessed. Another problem was due to the potentially unrealistic assumption that there is a one-to-one correspondence between PET and CT delineations. For example, lesions may have smaller uptake regions (on PET images) compared to outlines of lesions in CT images because of functional or metabolic characteristics of the tumor. In this study, we consider these issues and propose a co-segmentation method which is driven by the uptake regions from PET in finding the correct anatomical boundaries in the corresponding CT images. Our algorithm also uses a novel automatic background/foreground seed localization technique to make the whole system fully automatic.

2   Methods

Graph theoretic segmentation methods represent space elements (*spels* for short) of an image as a graph with spels as its nodes and *edges* defined by spel adjacency with cost values assigned to edges. These methods partition the nodes into two disjoint subsets representing the object and background. This process can be accomplished by finding the minimum cost/energy among all possible cut scenarios in the graph (as in graph-cut algorithms) or optimizing some sort of discrete energies combining boundary regularization with regularization of regional properties of segments [2]. However, a common problem with these is the "small cut" behavior. As a possible solution to this behavior, the random walk algorithm is more efficient than the conventional graph-cut algorithms in terms of handling ambiguities among object boundaries (i.e., weak edges among objects) and more accurate segmentations in noisy and low contrast images [5]. Since PET images are low resolution, and weak boundaries often exist in CT images, random walk segmentation is a natural choice for the simultaneous segmentation of PET and CT images. We *reformulate* random walk as a co-segmentation algorithm for delineating PET and CT images simultaneously and providing globally optimum segmentation results. To create a fully automated framework, we propose an automatic seed localization system by identifying *interesting uptake regions* (IUR)

from PET images, using these regions to identify foreground and background seeds, and propagating the detected background and foreground seeds to the corresponding CT images.

2.1   Automated Random Walk Co-segmentation

Let a connected and undirected graph G be represented as a pair $G = (V, E)$ with vertices/nodes $v \in V$ and edges $e \in E \subseteq V$ x V. Conventionally, a node v_i is said to be a neighbor of another node v_j if they are connected by an edge e_{ij}, which is weighted by w_{ij}. Since the graph is assumed to be connected and undirected: $w_{ij} = w_{ji}$. By following the recommendation in [5], we construct the weighting functions for PET and CT image modalities separately as

$$w_{ij}^{PET} = exp(-\beta^{PET}(I_i^{PET} - I_j^{PET})^2), \tag{1}$$

$$w_{ij}^{CT} = exp(-\beta^{CT}(I_i^{CT} - I_j^{CT})^2). \tag{2}$$

where I_i indicates the intensity at spel i, and β represents a weighting factor. Note that PET and CT images are obtained from the same scanner in the same scanning session; therefore they are registered so that there is a one-to-one *spel correspondence* between them. Conventionally, the desired random walker probabilities have the same solution as the combinatorial Drichlet problem [6]: $D[x] = \dfrac{1}{2}x^T L x$, where x denotes the probability (potential) assumed at each node [5], and L represents combinatorial Laplacian matrix. For each of the PET and CT modalities, this matrix can be formulated as:

$$L_{ij}^{\mathcal{X}} = \begin{cases} d_i^{\mathcal{X}} & \text{if } i = j \\ -w_{ij}^{\mathcal{X}} & \text{if } v_i^{\mathcal{X}} \text{ and } v_j^{\mathcal{X}} \text{ are adjacent nodes} \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

where $\mathcal{X}$ is either PET or CT, and d_i is the degree of a vertex for all edges e_{ij} incident on v_i and is defined as: $d_i = \sum_{e_{ij} \in E} w(e_{ij})$. Moreover, v^{PET} and v^{CT} are the nodes pertaining to the graph constructed on PET and CT images, respectively.

Simultaneous segmentation of PET and CT images on the graph requires a special representation of both data without losing information. From graph theory, it has been well known that given two graphs and their product as an outcome graph, an edge exists in the product graph if and only if an edge exists in both graphs [6]. Defining a special graph combining these two graphs, or a hypergraph, is a natural choice to satisfy this property. Given two graphs $G^{CT} = (V^{CT}, E^{CT})$ and $G^{PET} = (V^{PET}, E^{PET})$, without loss of generality, we define our special product graph as $G^{fuse} = (V^{fuse}, E^{fuse})$. Note that V^{CT} and V^{PET} have the same number of spels due to one-to-one spell correspondence. G^{fuse} has an important property that an edge exists in E^{fuse} *if and only if* the corresponding nodes are adjacent in both G^{CT} and G^{PET}. Thus,

$$V^{fuse} = \{(v_i^{CT}, v_i^{PET}) : v_i^{CT} \in V^{CT} \wedge v_i^{PET} \in V^{PET}\},$$
$$E^{fuse} = \{((v_i^{CT}, v_i^{PET}), (v_j^{CT}, v_j^{PET})) : \qquad (4)$$
$$(v_i^{CT}, v_j^{CT}) \in E^{CT} \wedge (v_i^{PET}, v_j^{PET}) \in E^{PET}\}.$$

We use the product graph G^{fuse} to segment objects simultaneously instead of using separate implementation of G^{CT} and G^{PET}. Since the critical points of $D[x]$ are assumed to be minima, finding these minima points yields the solution for the random walk probabilities. This requires an updated definition of the combinatorial Laplacian matrix on the product graph. Namely, the combinatorial Laplacian matrix (L^{fuse}) of the product graph G^{fuse} is defined as $L^{fuse} = (L^{CT})^{\alpha} \otimes (L^{PET})^{\theta}$, for some constants $0 \leq \alpha, \theta \leq 1$. Meanwhile, the initial probability distribution x^{fuse} of the product graph is $x^{fuse} = (x^{CT})^{\zeta} \otimes (x^{PET})^{\eta}$, where ζ and η are used to optimize the initial probability distributions subject to the constraint $0 \leq \zeta, \eta \leq 1$, and x^{CT} and x^{PET} denote initial probability distributions (i.e., priors) over nodes of G^{CT} and G^{PET}. Performing a random walk on the product graph G^{fuse} is equivalent to performing a simultaneous random walk on the graphs G^{CT} and G^{PET} [6]. Therefore, the combinatorial formulation of Drichlet integral can be re-written as

$$D[x^{fuse}] = \frac{1}{2}(x^{fuse})^T L^{fuse} x^{fuse} = \frac{1}{2} \sum_{e_{ij} \in E^{fuse}} w_{ij}^{fuse}(x_i^{fuse} - x_j^{fuse})^2, \qquad (5)$$

where a combinatorial harmonic function of x^{fuse}, satisfying the Laplace equation $\triangledown^2 x^{fuse} = 0$, minimizes Eqn. 5. We can decompose Eqn. 5 by considering prior probabilities and Laplacian matrices of labeled and unlabeled nodes separately as

$$D[x_u^{fuse}] = \frac{1}{2}[(x^{fuse})_l^T (x^{fuse})_u^T]\begin{bmatrix} L_l^{fuse} & B \\ B^T & L_u^{fuse} \end{bmatrix}\begin{bmatrix} x_l^{fuse} \\ x_u^{fuse} \end{bmatrix} \qquad (6)$$

where B corresponds to the sub-matrix in the matrix decomposition of L^{fuse}. Given the fact that the combinatorial Laplacian matrix L^{fuse} is positive semi-definite, critical points of $D[x^{fuse}]$ are only the minima, hence, differentiating $D[x^{fuse}]$ with respect to prior probability distributions of unlabeled nodes x_u^{fuse} and finding the minima yields $L_u^{fuse} x_u^{fuse} = -B^T x_l^{fuse}$, where L_u^{fuse} and B are known, and x_l^{fuse} is the prior for labeled node. Solving this equation for every x_u^{fuse} completes the binary labeling problem of co-segmentation on the graph G^{fuse}.

2.2 Automated Seed Localization

The goal in seed localization is to define the foreground and background regions in both PET and CT images. An overview of the proposed automated background and foreground seed localization is sketched in Fig. 1(f-i). Briefly, we
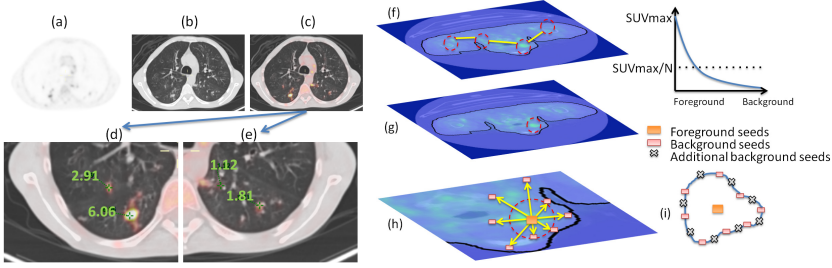
Fig. 1. Even though intensity profiles in CT images (b) show similar characteristics, SUV from PET images (a) might further characterize the nodules. Radioactivity uptake regions are shown in small nodules in (a and c)(and details of fused image (c) are shown in d and e). The concepts of interesting uptake region (IUR) detection and background/foreground seed localization are sketched in (f-i).

partition the image into a set of regions such that some of those regions (i.e., IURs) are more similar to each other than to those of other regions. We accomplish this selection procedure by defining an encoder function $c(.)$, which is nothing but a threshold interval for PET images:

$$c(\lambda) = \begin{cases} 1, \lambda \in [SUV_{max}^{global}/N, SUV_{max}^{global}], \\ 0, \qquad\qquad\qquad \text{otherwise}, \end{cases} \qquad (7)$$

where $(N > 1) \in \mathcal{R}$ is free parameter, and each region identified by the encoding function is considered as IUR. Once IURs are identified, we set the number of disconnected IURs as a hard constraint (i.e., number of objects to be segmented) on seed localizations. The seed localization procedure is as follows: (1) Both CT and PET images are median filtered to smooth the images. (2) We find the skin boundary from CT scan by using simple mathematical morphology (i.e., a few times opening followed by closing) and incorporate this information into the corresponding PET image (black outlines in Fig. 1f). (3) We find the IURs inside the body region and pertaining to the interval of $[SUV_{max}^{global}/N, SUV_{max}^{global}]$ by using Eqn. 7 (Fig. 1f). (4) For each IUR, the spels with the maximum SUVs (SUV_{max}^{local}) of that particular IUR are marked as foreground seeds (Fig. 1g). Note that maximum SUV of one particular IUR (i.e., SUV_{max}^{local}) does not necessarily equal SUV_{max}^{global}. (5) At each spel, marked as a foreground seed, we explore its neighborhood through an 8-connectivity graph labeling algorithm [6]. For all 8-directions starting from each foreground seed, we find locations of the very first spels with values less than or equal to the SUV_{max}^{global}/N. Those spels are marked as background seeds. Fig. 1h and i illustrate this procedure for a particular foreground seed. (6) We add additional background seeds into the spels lying in the spline connecting background seeds determined in the previous step (Fig. 1i). Fig. 2 shows automatically located seeds with (b) and without (a) additional background seeds. We find step 6 necessary to avoid any leakage in delineation of the abnormal anatomy in CT images. Fig. 2c shows an example of
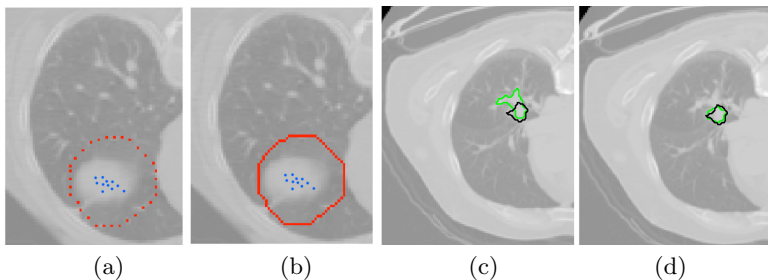
Fig. 2. (a) Automatically located background (red) and foreground (blue) seeds. (b) Additional background seeds are obtained by connecting initial background seeds using b-splines. (c) Ground truth (black) and random walk segmentations using only a limited number of background seeds. (d) The effect of the proposed seed localization method (with additional background seeds) in avoiding possible leakages.

the leakage occurring due to the close proximity of normal and abnormal tissues with similar intensity profiles while the effect of having additional background seeds on segmentation is shown in Fig. 2d.

3   Results

Data and Evaluation Metrics: A retrospective study involving 15 patients with PET-CT scans was performed. The resolution of PET scans is limited to spels of size 4mm x 4mm x 4mm. The patients considered have infectious lung disease abnormality patterns including ground glass opacities, consolidations, nodules, tree-in-bud, lung tumors, and non-specific lung lesions. While PET scans consist of more than 300 slices per patient, CT scans have the same number of slices but different in-plane resolution with 0.98 mm x 0.98 mm x 1.5 mm spel size. Dice similarity coefficients (DSC) and Hausdorff distance (HD) are used to evaluate the accuracy of segmentations. High DSC and low HD values indicate goodness of the image segmentation method. The ground truth segmentations were obtained from manual delineations of two expert interpreters on PET-CT data, and average DSC values were reported.

Qualitative and Quantitative Evaluation: Fig. 3 shows segmentation examples. Co-segmentation of PET-CT images (blue), segmentation of CT images (green), and PET images (yellow) are shown in first, second, and third columns, respectively. The fourth column reveals all segmentations overlaid together for comparison. In all images, ground truth segmentations are shown in black. It can be seen that the co-segmentation results are superior to delineation using only CT and only PET images, and agree well with the ground truth delineations. Fig. 4 reports the average DSCs and HDs of delineations over 15 subjects and comparison to the method presented in [2]. Note that co-segmentation on PET-CT images is superior in accuracy of segmentations to other methods.
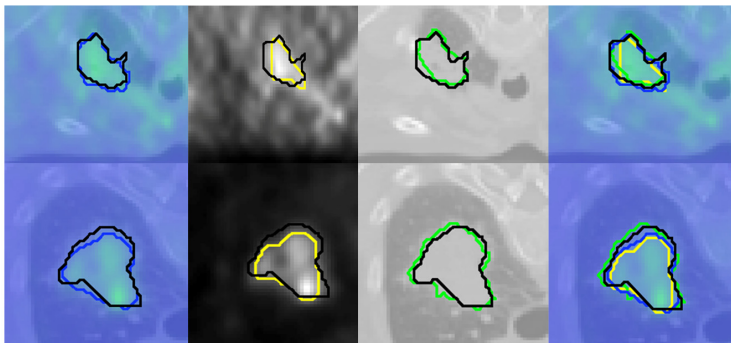
Fig. 3. Two different segmentation examples of uptake regions are shown in each column. First column: co-segmentation (blue) and ground truth (black) are overlaid. Second column: ground truth (black) and segmentation from PET only (yellow). Third column: ground truth (black) and segmentation from CT only (green). Fourth column: all segmentations and ground truth are overlaid together.
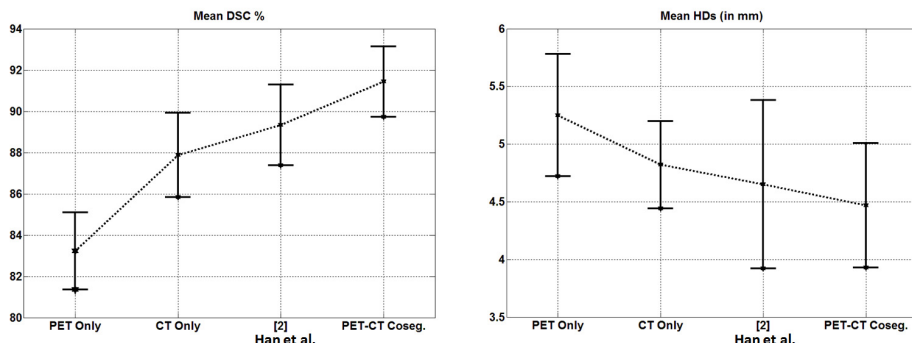
Fig. 4. Mean DSCs and HDs are enlisted. DSC ratios: PET Only: 83.23 ∓ 1.87, CT Only: 87.88 ∓ 2.04, Han et al.: 89.34 ∓ 1.95, PET-CT cosegm.: 91.44 ∓ 1.71. HDs ratios: PET Only: 5.25 ∓ 0.53, CT Only: 4.82 ∓ 0.38, Han et al: 4.65 ∓ 0.73, PET-CT cosegm.:4.47 ∓ 0.54.

Training, Parameter Selection, and Computational Issues: First, we upsampled the PET images so that the PET and CT images have the same size, and each spel in PET has its correspondence in CT in the same spatial coordinate. We used a training set consisting of PET-CT scans of 5 patients (different from the test data set) to train parameters of the segmentation process explained below. In defining the product graph, we weight the corresponding combinatorial Laplacian matrices (L^{CT} and L^{PET}) with α and θ and the initial probability distributions x^{CT} and x^{PET} with ζ and η, respectively. We noted that multiplication and summation of Laplacian matrices do not have significant effects on the segmentation results if ζ and η are set to 1. The best possible

combination of (α, θ) is found to be $(0.2, 0.8)$ when $\zeta = \eta = 1$. Other parameters were set to: $\sigma^{PET} = \sigma^{CT} = 1$, $\beta^{PET} = 0.3$, and $\beta^{CT} = 0.7$. Lastly, based on our experimental results and empirical observations, we set a slightly tighter bound on SUV_{max}/N by setting $N = 2$ compared to the conventional clinical usage where N is usually 2.5. The combined running time for detecting all seeds and delineating the corresponding binary labeling problem per slice does not exceed 2 seconds at maximum, and averages 0.8 seconds. The maximum number of slices showing high tissue activity does not exceed 50, therefore all detection and segmentation procedures end within an average of 40 seconds (using an Intel (R) workstation with Core(TM) i7 CPU 930 running at 2.80 GHz with 12 GB RAM). The proposed algorithm can be run in either pseudo-3D or 3D directly.

4   Discussion and Conclusion

We proposed a joint-segmentation framework for anatomical and functional images. Our approach differs from the approach reported in [2] in the following manners: (1) We propose a completely automated method of segmentation, therefore user-interaction is not required. As a result, the proposed method is faster and reproducible. (2) We do not have constraints about one-to-one correspondence of abnormal anatomy in PET and CT images, as it is possible that a tumor can reveal increased radioactivity uptake only in some areas inside the tumor region, or it may be possible for an uptake region to enclose both normal and abnormal anatomy in CT scans. Instead, it is semantically more meaningful and reliable to drive the segmentation of both images based on the guidance of PET images. (3) Although MRF based segmentation of images on graphs is shown to be useful in many applications, incorporating an additional energy term, similar to the one proposed in [2], may not be the best solution for avoiding the leakage in CT and PET delineations. In contrast, the proposed method is more powerful in terms of handling image noise and low contrast and therefore more suited to PET-CT segmentations. Another advantage of our method is the special graph model constructed by a product graph via taking into account the anatomical and functional image features to derive algorithms for accurately finding the most likely IUR boundaries. Initial results on 15 clinical PET-CT images (i.e., thousands slices) show high accuracy in delineations by the proposed method. Furthermore, the proposed method considerably reduces the time for manual segmentations. As an extension of this work, we aim to generalize the proposed co-segmentation method for MRI-PET and multi-fusion MRI-CT-PET segmentations.

References

[1] Judenhofer, M.S., et al.: Simultaneous PET-MRI: a new approach for functional and morphological imaging. Nature Medicine 14, 459–465 (2008)
[2] Han, D., Bayouth, J., Song, Q., Taurani, A., Sonka, M., Buatti, J., Wu, X.: Globally Optimal Tumor Segmentation in PET-CT Images: A Graph-Based Co-segmentation Method. In: Székely, G., Hahn, H.K. (eds.) IPMI 2011. LNCS, vol. 6801, pp. 245–256. Springer, Heidelberg (2011)

[3] Montgomery, D.W., et al.: Fully automated segmentation of oncological PET volumes using a combined multiscale and statistical model. Medical Physics 34(2), 722–736 (2007)

[4] Jentzen, W., et al.: Segmentation of PET Volumes by Iterative Image Thresholding. J. Nucl. Med. 48, 108–114 (2007)

[5] Grady, L.: Random Walks for Image Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(11), 1768–1783 (2006)

[6] Harary, F.: Graph Theory. ABP Publishing (1994)

Using Multiparametric Data with Missing Features for Learning Patterns of Pathology

Madhura Ingalhalikar[1], William A. Parker[1], Luke Bloy[1],
Timothy P.L. Roberts[2], and Ragini Verma[1]

[1] Section of Biomedical Image Analysis, University of Pennsylvania,
Philadelphia, PA, USA
{Madhura.Ingalhalikar,Ragini.Verma}@uphs.upenn.edu
[2] Lurie Family Foundations MEG Imaging Center, Department of Radiology,
Children's Hospital of Philadelphia, Philadelphia, PA, USA⋆

Abstract. The paper presents a method for learning multimodal classifiers from datasets in which not all subjects have data from all modalities. Usually, subjects with a severe form of pathology are the ones failing to satisfactorily complete the study, especially when it consists of multiple imaging modalities. A classifier capable of handling subjects with unequal numbers of modalities prevents discarding any subjects, as is traditionally done, thereby broadening the scope of the classifier to more severe pathology. It also allows design of the classifier to include as much of the available information as possible and facilitates testing of subjects with missing modalities over the constructed classifier. The presented method employs an ensemble based approach where several subsets of complete data are formed and trained using individual classifiers. The output from these classifiers is fused using a weighted aggregation step giving an optimal probabilistic score for each subject. The method is applied to a spatio-temporal dataset for autism spectrum disorders (ASD)(96 patients with ASD and 42 typically developing controls) that consists of functional features from magnetoencephalography (MEG) and structural connectivity features from diffusion tensor imaging (DTI). A clear distinction between ASD and controls is obtained with an average 5-fold accuracy of 83.3% and testing accuracy of 88.4%. The fusion classifier performance is superior to the classification achieved using single modalities as well as multimodal classifier using only complete data (78.3%). The presented multimodal classifier framework is applicable to all modality combinations.

1 Introduction

Pattern classification techniques are generating increasing interest in the neuroimaging community as they are powerful in learning the patterns of pathology

---

from a population, assign a probabilistic score to each subject which characterizes pathology on an individual basis and aid in assessing treatment in conjunction with other clinical scores [1,2]. The earlier single modality [1,2] studies have given way to multimodality classifiers that can potentially aid in exploring additional dimensions of pathology patterns and provide a rich multiparametric signature or profile with increased diagnostic accuracy [3,4]. However, none of these studies account for a challenging problem plaguing clinical studies, that data from some modalities could be missing, as the subjects do not complete the entire study due to enhanced pathological severity or scanner issues and noise that force the studies to partially discard the data. Removing subjects with incomplete datasets from the study (as is the approach adopted by the traditional multimodal classification studies) reduces the already small sample size and diminishes the information content in the dataset, thus making the classifier decision unreliable since it does not account for the pathology patterns from the subjects who were unable to complete the clinical study due to their more severe form of pathology. Further, it limits the dimensionality of the multimodal approach since the probability of a subject being excluded increases with the number of modalities attempted.

In statistical theory, missing value problems are addressed using various strategies based on the patterns in missing data [5]. For randomly missing data, imputation techniques that substitute or fill in the missing items are commonly used [5,6]. Imputation methods can include substitutions like mean or median of the feature or multiple substitutions which replace each missing value with a set of plausible ones, reflecting the underlying uncertainty in the data [6]. Thus, the multiple imputation technique is considered as one of the effective methods to handle partial data. Other well-established strategies to deal with missing data involve model based procedures like expectation maximization (EM) which potentially recover unknown values from similar samples. Finally, simple decision tree classifiers have also been utilized as they avoid the missing data completely [6].

However, most of the above methods perform substitution in some way or the other, which usually interpolates and may create spurious data and thus cannot be completely trusted when the percentage of missing data is high ($\approx 30\%$ and above). Moreover, if the missing data is associated with an extreme in the pathologic condition, which is widely true, interpolation would become extrapolation making the data highly unreliable. Finally, these methods just attempt to fill in and thus do not directly take the classification problem into consideration [7]. Recent machine learning literature has shown ensemble classifiers to be an effective way to accommodate sparse data by using weak classifiers and then boosting the performance by combining the output [7], [8].

In this paper we present an ensemble based classification framework that has the potential to handle spatio-temporal multimodal data with a high percentage of missing values in a petite sample size. The method learns patterns of pathology on different subsets created from the original data and aggregates the output from all the classifiers using a weighting strategy, giving an optimal probabilistic score for each subject. The method considers subjects with complete or incomplete data

in training the classifier as well as in testing, without filling in the missing values. We apply this method on a population with Autism Spectrum Disorder (ASD) and typically developing controls (TDC), where the pathology can be investigated by creating spatio-temporal classifiers that utilize the spatial features from diffusion tensor imaging (DTI) of white matter and the temporal features computed from magnetoencephalography (MEG) data increasing the classification accuracy over the single modality classifiers.

2   Methods

We use ensemble classifiers, i.e. a pool of classifiers, each trained using a subset of the original dataset [7]. The output from all the classifiers is fused together to boost the overall performance. This fusion is a weighted aggregation based on the classification accuracy of each classifier as well as the similarity between the features used in the training and the testing set. We successfully demonstrate the applicability of this method on a spatio-temporal dataset derived from MEG features and DTI features.

2.1   Classification of a Dataset with Missing Values

Consider a dataset with n subjects $(x_1, x_2, .., x_n)$ and m features. Each subset (S) is defined by a collection of subjects $(< n)$ that have complete data for a specific set of features $s(s \leq m)$. Subsets are formed such that the total number of subsets t encompass all the subjects and the features from the original dataset. The decision of how many subsets to create depends upon what features are missing in the training and testing samples. In the case where all the features have some part missing, $2^m - 1$ subsets can be created.

A classifier model (e.g. LDA, SVM etc) is trained over each subset resulting in t outputs. Thus, for a subject x_i, the classification output can be formulated as $(O(x_i, S_1), .., O(x_i, S_t))$ resulting from t subsets. At the next stage, all the outputs from individual classifiers are combined to boost the overall performance. While fusing the outputs it is important to note that some subsets could be more valuable than others, depending on the subject under testing. Therefore, in the aggregation stage the final output for a subject is given by the weighted combination of the subset outputs [7]. For example, for a subject x_{test} under testing the final output is given by 1.

$$O_{agg}(x_{test}) = \frac{1}{\sum_i \frac{1}{\varphi_{i,x_{test}}}} \sum_{i=1}^{t} \frac{1}{\varphi_{i,x_{test}}} O(x_{test}, S_i) \tag{1}$$

In 1, $\varphi_{i,x_{test}}$ is the expected error of the classifier (given by equation 2) for subset S_i which depends on the general accuracy of the classifier for S_i as well as the number of features used in the classifier.

$$\varphi_{i,x_{test}} = \frac{1}{\eta_{x_{test}}} \sum_{x \in training} d(x, x_{test}) \varphi_{i,x} \tag{2}$$

where

$$d(x, x_{test}) = \sqrt{\|x - x_{test}\|^2 f(s_x^T K s_{x_{test}})} \tag{3}$$

$$\eta_{x_{test}} = \sum_{x \in training} d(x, x_{test}) \tag{4}$$

$$\varphi_{i,x} = (O(x, S_i) - Y(x))^2 f(s^T K s_x) \tag{5}$$

The training accuracy of each classifier is accounted via $\varphi_{i,x}$ in equation 2. $\varphi_{i,x}$ is the expected error for a subject under training given by equation 5, where $Y(x)$ is the known label and $O(x, S_i)$ is the output of subset S_i for a training subject x. The $d(x, x_{test})$ term in equation 2 takes into account the distance between two samples. The similarity between the feature space of the subset and the subject is given by function f in equations 3 and 5 where K is a diagonal matrix that weighs the features based on their information content [7]. Here we define f as $f(v) = 1/v$, a non-increasing function, accounting for the similarity term as well as the feature ranking.

2.2   Classification Using MEG+DTI Features

Feature Extraction. In this study, we consider 3 categories of features. These features are associated primarily with language impairment in ASD, but the framework is applicable to any set of MEG-DTI features. The features used in this study are: (i) the latency of auditory evoked neuromagnetic field 100ms component called M100 [9], (ii) the latency of magnetic mismatch field (MMF), which is a response component reflecting detection of 'change' in the auditory stream [10], and (iii) Fractional anisotropy (FA) and mean diffusivity (MD) measures from 37 ROI's created by a normalized cuts clustering method [11] in WM areas of brain associated with language (figure 1), providing 74 values.

Feature Ranking in DTI. In the subsets that contain DTI as a feature, we perform feature ranking on the 74 DTI attributes. This step provides us with attributes that most contribute to the patient-control classification and aids in minimizing the classification error. To find a compact discriminatory subset of features, we choose a ranking and selection method known as the signal-to-noise (s2n) ratio coefficient filter [12]. For the jth DTI attribute vector dti_j and class labels Y, the signal to noise ratio is given by equation 6. In this equation, $\mu(y+)$ and $\mu(y-)$ are the mean values while $\sigma(y+)$ and $\sigma(y-)$ are the variances for class $y+$ and $y-$ respectively. Based on these s2n coefficients, the features are ranked and a subset of the top ranked features is implemented in the classification.

$$s2n(dti_j, Y) = \frac{\mu(y+) - \mu(y-)}{\sigma(y+) + \sigma(y-)} \tag{6}$$

Training, Cross Validation and Testing. We have created a generalized framework; therefore, any kind of classifier (e.g. SVM, QDA etc.) would work in

this ensemble setup. We implement a simple linear discriminant analysis (LDA) classifier on each of our subsets as the aim here is to demonstrate that incorporating missing data from our multimodal features aids in classification.

Since our missing data is spread over all three features, we use total of 7 subset combinations (M100, MMF, DTI, M100+DTI, MMF+DTI, MMF+M100 and MMF+M100+DTI). If a subset contains DTI as a feature, then s2n ranking is performed only on the DTI features. The number of top ranked DTI features to be retained in the classification process is based on the minimum cross validation error computed from the error plot that is constructed by using different number of features in the cross validation [12]. The ranking matrix K defined in section 2.1 is set to identity, implying that M100, MMF and DTI provide equal information.

We compute a probabilistic abnormality score for each subject using a 5-fold cross validation on the training data which is permuted 100 times for generalization of the folds. Finally, the trained classifier framework is applied to test data with missing values.

3   Results

Dataset and Preprocessing. Our dataset consisted of 138 subjects (42 TD and 96 ASD), out of which 55 subjects had complete data while others had some feature missing (60.1% subjects with partial data). 30% of subjects were missing MMF, 15.7% were missing M100 and 38% were missing DTI. We randomly picked 112 subjects (51 complete and 61 subjects with partial data, making 54.4% missing data) for training and the other 26 (4 complete and 22 partial data) as test data.

The MEG recordings were performed using a CTF 275-channel biomagnetometer with the following protocol: (i) binaural auditory presentation of brief sinusoidal tone stimuli at 45dB SL. M100 latency was determined from the source modeled peak of the stimulus-locked average of 100 trials of each token (ii) binaural auditory presentation of interleaved standard and deviant tone and vowel tokens (/a/, /u/). Mismatch field (MMF) latency was determined from the subtraction of superior temporal gyrus (STG) source-modeled responses for each token as deviant vs. standard. The DTI data were acquired on Siemens 3T
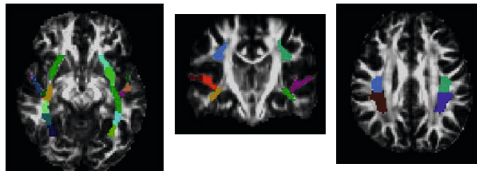
Fig. 1. Figure displays the ROI's in which mean FA and MD were used in the classifier. These ROI's were computed using normalized cuts algorithm in the areas associated with language.
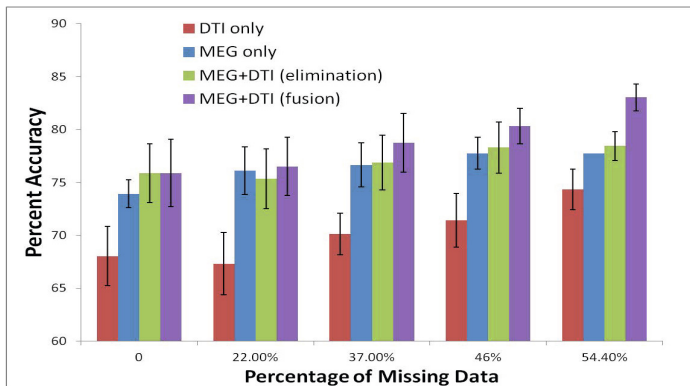
Fig. 2. Bar chart comparing the accuracies of 4 different classifiers. The performance of the fusion method becomes superior with more missing data. Multimodal classifiers perform better than single modality classifiers in all the cases. The x-axis displays the percentage of missing data starting with 51 complete subjects and then randomly adding 15 subjects with missing data in each case until all the subjects with missing data were added (total 112 subjects) in the last case.

Verio^{TM} scanner using the Stejkal Tanner diffusion weighted imaging sequence (2mm isotropic resolution) with b=1000 mm/s^2 and 30 gradient directions.

The mean scalar features (FA and MD) from DTI were computed in each of the 37 ROIs (figure 1). For this initially, language related WM ROI's (superior temporal white matter (STWM), inferior and superior longitudinal fasciculi (SLF, ILF), inferior fronto-ocipital fasciculus (IFOF)) were derived from a standard atlas called EVE [13]. The size of these ROI's is large enough to smooth out the effect of mean FA and MD. Therefore, we implemented a normalized cuts algorithm [11] that was based on a variance threshold computed over the DTI images to divide these ROI's into smaller regions with homogeneous WM.

Training, Cross Validation and Testing. Using 5-fold cross validation, we computed the classifier score for each subject in training. For our dataset, the top 15% DTI features were employed based on their optimal performance in the cross-validation of DTI classifier. Figure 2 shows a bar chart of training accuracies for classifiers with only DTI, only MEG (MMF and M100), DTI and MEG with deletion of incomplete data and DTI and MEG with the fusion approach. When only the subjects with complete data (N=51) were considered, the classifier performed with only 78.3% accuracy, while when using the whole sample (N=112), including those with missing data, the accuracy increased to 83.3%.

Out of the 7 subset classifiers, other than the classifier with all 3 features, the individual MMF classifier (77.7%) and the MMF+DTI classifier (79.4%) added to the overall discriminative power. Individual DTI classifier performed with an accuracy of ≈75%, higher than the individual M100 classifier (≈70%), suggesting that combining modalities performed better (note that the number of subjects in

<center>(a)   (b)</center>
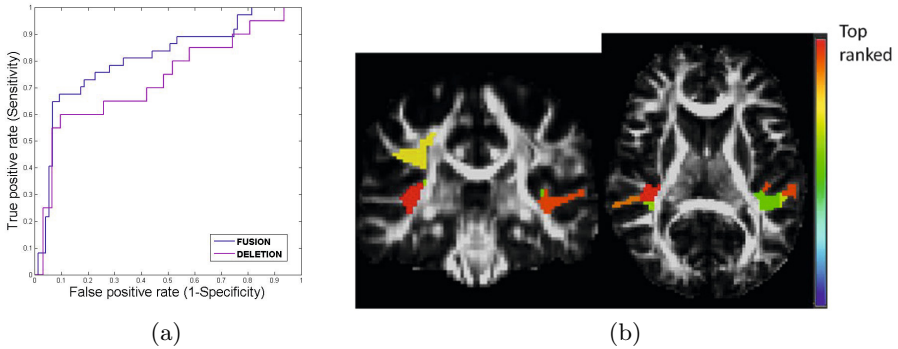
Fig. 3. (a) ROC curve for classifier with complete data (51 subjects) and ROC for classifier with 54.4% missing data (112 subjects). (b) ROI's that we frequently selected by the DTI feature selection technique in the 5-fold classification.

each classifier were different). The ensemble framework gains from the diversity of the subset classifiers, boosting the overall performance.

Figure 3(a) displays the receiver operating characteristic (ROC) curves for the cases with complete data (51 subjects) and missing data based (112 subjects) on the 5-fold validation. The area under curve for the first case was 0.73 that increased to 0.82 in the second case. Finally, we classified the 26 subjects reserved as testing samples, on the fusion classifier trained on the 112 subjects. The testing accuracy was 88.5% where 100% a (4/4) on subjects with complete data and 86.4% (19/22) subjects with partial/missing data were classified correctly.

All the brain regions employed to extract DTI features are specific to language impairment in ASD. The feature selection ranks these regions in each of the cross-validation loop suggesting that the selected regions have more discriminative power. The most frequently selected features (in the 5-fold validation) via s2n ranking scheme are shown in figure 3(b). These mainly include right SLF and right and left STWM. Other regions like left SLF were also involved in the 5-fold classification, but were not selected very frequently.

4   Conclusion

We have presented a classification technique that can build classifiers on multimodal data with missing modalities/features. We applied it to a problem involving spatio-temporal (MEG and DTI) data with features that were associated with language abnormalities in ASD. When multiple modalities are utilized within a classification framework, the aggregate output for each subject can potentially define a more comprehensive quantification of pathology than when used individually. Such an approach necessitates a means for handling subjects with incomplete data, such as presented here. Our ensemble approach demonstrated superior performance when compared with those utilizing the smaller complete samples, suggesting that utilizing the subjects with incomplete data was advantageous to

correctly learning the patterns of difference. The internal DTI feature ranking pointed out the regions that were responsible for language impairment in ASD while the subset classifiers in the ensemble provided insight into the relative contributions of DTI and MEG to classification.

In large studies that involve multimodality imaging data together with psychological scores, genomic data etc.,a high percentage of missing data is expected. Our generalized framework can be readily applied to such problems and will have a high impact.

References

1. Fan, Y., Shen, D., Gur, R.C., Gur, R.E., Davatzikos, C.: Compare: classification of morphological patterns using adaptive regional elements. IEEE Trans. Med. Imaging 26(1), 93–105 (2007)
2. Ingalhalikar, M., Parker, D., Bloy, L., Roberts, T.P.L., Verma, R.: Diffusion based abnormality markers of pathology: toward learned diagnostic prediction of asd. Neuroimage 57(3), 918–927 (2011)
3. Batmanghelich, N., Dong, A., Taskar, B., Davatzikos, C.: Regularized Tensor Factorization for Multi-Modality Medical Image Classification. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part III. LNCS, vol. 6893, pp. 17–24. Springer, Heidelberg (2011)
4. Zhang, D., Wang, Y., Zhou, L., Yuan, H., Shen, D., Initiative, A.D.N.: Multimodal classification of alzheimer's disease and mild cognitive impairment. Neuroimage 55(3), 856–867 (2011)
5. Little, R., Rubin, D.: Statistical Analysis with Missing Data. John Wiley (2002)
6. Garcia-Laencina, P.J., Sancho-Gomez, J.L., Figueiras-Vidal, A.R.: Pattern classification with missing data: a review. Neural Comput. and Application 19(2), 263–282 (2009)
7. Ghannad-Rezaie, M., Soltanian-Zadeh, H., Ying, H., Dong, M.: Selection-fusion approach for classification of datasets with missing values. Pattern Recognit. 43(6), 2340–2350 (2010)
8. Wang, C., Liao, X., Carin, L.: Classification of incomplete data using dirichlet process priors. Journal of Machine Learning Research 11, 3269–3311 (2010)
9. Cardy, J.E.O., Flagg, E.J., Roberts, W., Brian, J., Roberts, T.P.L.: Magnetoencephalography identifies rapid temporal processing deficit in autism and language impairment. Neuroreport 16(4), 329–332 (2005)
10. Roberts, T.P.L., et al.: Meg detection of delayed auditory evoked responses in autism spectrum disorders: towards an imaging biomarker for autism. Autism Res. 3(1), 8–18 (2010)
11. Bloy, L., Ingalhalikar, M., Eavani, H., Schultz, R.T., Roberts, T.P.L., Verma, R.: White matter atlas generation using hardi based automated parcellation. Neuroimage (August 2011)
12. Guyon, I., Elisseeff, A.: An introduction to variable and feature selection. J. Mach. Learn. Res. 3, 1157–1182 (2003)
13. Oishi, K., et al.: Atlas-based whole brain white matter analysis using large deformation diffeomorphic metric mapping: application to normal elderly and alzheimer's disease participants. Neuroimage 46(2), 486–499 (2009)

Non-local Robust Detection of DTI White Matter Differences with Small Databases

Olivier Commowick and Aymeric Stamm

VISAGES: INSERM U746 - CNRS UMR6074 - INRIA - Univ. of Rennes I, France
Olivier.Commowick@inria.fr

Abstract. Diffusion imaging, through the study of water diffusion, allows for the characterization of brain white matter, both at the population and individual level. In recent years, it has been employed to detect brain abnormalities in patients suffering from a disease, e.g. from multiple sclerosis (MS). State-of-the-art methods usually utilize a database of matched (age, sex, ...) controls, registered onto a template, to test for differences in the patient white matter. Such approaches however suffer from two main drawbacks. First, registration algorithms are prone to local errors, thereby degrading the comparison results. Second, the database needs to be large enough to obtain reliable results. However, in medical imaging, such large databases are hardly available. In this paper, we propose a new method that addresses these two issues. It relies on the search for samples in a local neighborhood of each pixel to increase the size of the database. Then, we propose a new test based on these samples to perform a voxelwise comparison of a patient image with respect to a population of controls. We demonstrate on simulated and real MS patient data how such a framework allows for an improved detection power and a better robustness and reproducibility, even with a small database.

1 Introduction

Diffusion weighted imaging is an MRI modality that provides information about water diffusion within tissues. It has therefore gained much interest for the study of brain white matter architecture. In particular, it may be utilized for the detection of structural differences related to a disease. Reported studies on diseases usually fall within two categories: (i) group comparisons between a population of healthy subjects and a group of patients suffering from the disease and (ii) comparison of one patient to a set of healthy controls. The former aims at characterizing the overall course of a disease while the latter focuses on detecting its early signs and, possibly, its future evolution.

Both approaches are of great interest to understand a disease. In this work, we are interested in diffusion imaging for multiple sclerosis (MS). MS is a demyelinating disease, causing both lesions visible on conventional MRI and diffuse damage to the brain white matter architecture that may be visible in diffusion imaging [1]. Having a robust detection of that diffuse damage for a specific patient is crucial as it could help to predict how the disease will evolve in time, and potentially allow to adapt the treatment.

Recent works on the comparison of diffusion images have first focused on scalar values extracted from the diffusion tensor, such as mean diffusivity (MD) or fractional anisotropy (FA). For example, Filippi et al. [2] presented a study on manually defined regions of interest demonstrating an MD increase and an FA decrease in specific regions for MS patients brains. However, utilizing only a scalar value may discard a large part of the tensor information and decrease the precision of the comparison. To overcome this problem, several groups have proposed methods to utilize the full tensor either for population comparison (Lepore et al. utilized the Hotelling's T^2 test on tensors for HIV patients [3], Whitcher et al. [4] the Cramers test on tensors), or for patient to group comparison (Commowick et al. [5]). These works have demonstrated that a test based on the full tensor information yields more precise comparisons. Finally, when high quality data is available (HARDI acquisitions), one may now consider higher order models such as orientation distribution functions (ODFs) to get improved sensitivity in crossing fibers regions where the diffusion tensor performs poorly [6].

Independently of their strengths and weaknesses, comparison methods usually rely on a parametric or permutation statistical test. Such approaches often require large databases either to ensure that the distribution of the test statistic matches the hypothesized one or to make the permutation test data independent. However, in medical imaging studies, databases are usually small due to the difficulty to recruit patients and volunteers, and they may be even smaller when parameters such as age or sex must match between the control database and the patients. In those cases, the chosen statistical test may become erroneous and generate either false positive or false negative detections.

In addition, all automatic approaches need a common reference frame that is often constructed from the healthy subjects by means of non linear registration (so called atlas construction [7]). However, such registration methods are not perfect and may be prone to errors due to noise and artifacts. Such errors may further corrupt the comparison performance.

To tackle these issues, we propose a new methodology for the robust detection of white matter differences at a patient level. It is based on ideas recently introduced for non-local means denoising [8] and segmentation [9], adapted for a patient-to-group comparison of diffusion models that can be represented as vectors in a vector space (e.g diffusion tensors or ODFs). We present in Section 2 the overall comparison method. We then apply this new method to simulated data and real data of multiple sclerosis patients demonstrating higher accuracy and reproducibility for differences detection over state-of-the-art methods.

2   Non-Local Means for the Comparison of Diffusion Data

In the following, we assume that a database of M images I_m has been constituted, i.e. all these images have been non linearly registered to a common reference system, and that we are interested in comparing voxel by voxel an image T to the reference database. We propose an algorithm relying on the non-local means framework [10] optimized by Coupé et al. for medical image denoising [8] and

segmentation [9]. For each point x of T, we define a patch $B(x)$ (half size h) around it and follow these main steps:

- For each image I_m, search for patches $B_m(x_j)$ similar to the patch $B(x)$ of T in a neighborhood $N(x)$ around x
- Associate a weight w_{mj} to each patch $B_m(x_j)$, depending of the similarity between $B(x)$ and $B_m(x_j)$ (Section 2.1)
- Keep the center voxel D_{mj} of each patch and associate it to its weight w_{mj}
- Utilize the set of weighted samples to perform the comparison between T and $I_m, m \in \{0, ..., M\}$ (Section 2.2)

This framework has several advantages: it may help to account for potential registration errors onto the common template for comparison, and it may significantly increase the number of samples to perform the voxelwise comparison even though the database consists of a limited number of images.

2.1   Similarity Weights between Patches

The selection of patches is a crucial point as it will define the relative importance of each patch in the final differences detection step. We consider that the model chosen to describe the diffusion of water molecules may be represented as a vector, e.g. tensors in the Log-Euclidean framework [11] or ODFs on a spherical harmonics basis [6]. Before comparing patches, a preselection is performed for speed reasons and to avoid the degeneracy of the patches weights w_{mj}:

1. Compute the Log-Euclidean distance between the covariance matrices $\Sigma_{B(x)}$ and $\Sigma_{B_m(x_j)}$: if it exceeds the average distance between any two covariance matrices $\Sigma_{B_m(x)}$ of the database, discard $B_m(x_j)$, otherwise proceed to the next step;
2. Compute Hotelling's T^2 statistic [12] to test for mean differences between $B(x)$ and $B_m(x_j)$ using the pooled covariance matrix: if it exceeds the average statistic computed from any two patches $B_m(x)$, discard patch $B_m(x_j)$.

For the remaining patches, we then compute their weights. The weight w_{mj} between two patches $B(x)$ and $B_m(x_j)$ is defined as a function of the sum of squared differences between the two patches:

$$w_{mj} = e^{-\frac{1}{2\beta|B(x)|} \sum_{i \in B(x)} \Delta_i^T \hat{\Sigma}^{-1}(x) \Delta_i} \tag{1}$$

where $\Delta_i = I_m(i + x_j - x) - T(i)$ are the differences between corresponding voxels of the patches, $|B(x)|$ is the number of voxels in $B(x)$, β a user-defined scale parameter and $\hat{\Sigma}(x)$ is the local noise covariance around x on T. These weights characterize the similarity between patches and vary between 0 and 1: 1 is reached when the two patches are equal, 0 corresponds to a total disagreement.

Since structures with different orientations may occur, the noise covariance $\hat{\Sigma}$ is estimated locally. Computing it globally over the whole image could indeed lead to an over-estimation and therefore to biased weights. Coupé et al. [8] proposed a method to estimate such a local noise variance on scalar valued

images. Here, we extend it to vector-valued images. $\hat{\Sigma}(x)$ is estimated from voxels in patch $B(x)$:

- For each voxel x_i in $B(x)$, consider a small neighborhood of N voxels around it (e.g. the 26 neighbors of x_i). Pseudo-residuals ϵ_{x_i} are computed as:

$$\epsilon_{x_i} = \sqrt{\frac{N}{N+1}} \left(T(x_i) - \frac{1}{N} \sum_{x_j \in N(x_i)} T(x_j) \right)$$

- The local image noise covariance is defined from these pseudo-residuals:

$$\hat{\Sigma}(x) = \frac{1}{|B(x)|} \sum_{x_i \in B(x)} \epsilon_{x_i} \epsilon_{x_i}^T$$

2.2   Comparison of Weighted Data Samples

We have constituted a list of weighted samples $S = \{S_1, \ldots, S_M\}$ at each voxel x of T, where $S_m = \{(D_{m0}, w_{m0}), \ldots, (D_{mJ}, w_{mJ})\}$. We now utilize these samples to confront the patient image to the healthy subjects database. We compute the weighted mean μ_x and weighted covariance matrix Σ_x at each point x as:

$$\mu_x = \frac{1}{\sum_{i,j} w_{ij}} \sum_{i,j} w_{ij} D_{ij} \tag{2}$$

$$\Sigma_x = \frac{\sum_{i,j} w_{ij}}{\left(\sum_{i,j} w_{ij} \right)^2 - \sum_{i,j} w_{ij}^2} \sum_{i,j} w_{ij} (D_{ij} - \mu)(D_{ij} - \mu)^T$$

These estimates are very interesting as they take into account the similarity of each patch in the estimation of the mean and covariance. We then test for voxelwise differences by computing the Mahalanobis distance at each point:

$$Z^2(D_x, \mu_x, \Sigma_x) = (D_x - \mu_x)^T \Sigma_x^{-1} (D_x - \mu_x) \tag{3}$$

where D_x is the vector value of the patient image at point x (e.g. log-tensor or ODF value). Considering there are enough samples, this squared distance follows a χ^2 distribution with d degrees of freedom, where d is the vector dimension, and a p-value is computed from Z^2 as:

$$p_x = 1 - F_{\chi_d^2} \left(Z^2(D_x, \mu_x, \Sigma_x) \right) \tag{4}$$

where $F_{\chi_d^2}$ is the cumulative distribution function of a χ^2 distribution with d degrees of freedom.

3   Results

Our method has two main parameters: the patch size and the search neighborhood. The smaller the sizes, the closer the method gets to [5]. On the contrary, large sizes tend to increase the number of false negatives. We fixed the parameters on the basis of qualitative results on several patients: patch size of $3 \times 3 \times 3$ ($h = 1$) and a neighborhood for patch search of 4 voxels in every direction. In addition, we have set β - Eq. (1) - to 1 as is suggested by Coupé et al. [8].

3.1   Experiments on Simulated Data

We first present a quantitative study on simulated images. Starting from a reference diffusion tensor image (Fig. 1.a), 90 images were simulated by adding Rician noise to the DWI. Then, a patient image was simulated by inserting lesions, i.e. tensors swollen in the two non principal directions. To illustrate the detection power of our method and its robustness to database size, we randomly selected from the 90 images subgroups of 15 to 90 images and used them as the reference database to compare to the simulated patient. Fig. 1 shows the average Dice score results of our method (M_2) and the one proposed in [5] (M_1).

This figure illustrates well the issues arising when using a small database for differences detection. As the sample size decreases, M_1 performs worse, mainly due to a large number of false positives being detected (see Fig. 1.c).

(a)   (b)

(c)   (d)   (e)
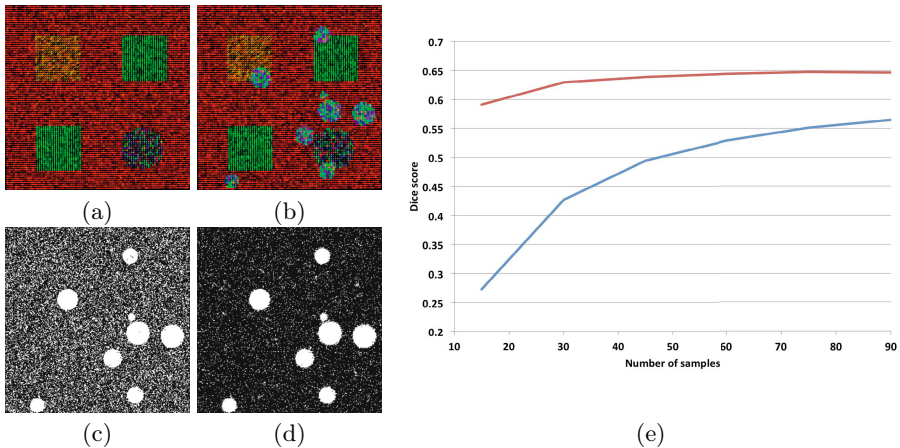
Fig. 1. Quantitative Detection Power on Simulated Data. Left: Illustration of one noisy reference database image (a) and the simulated lesions image (b), as well as results of detection utilizing 15 images from the database with M_1 (c) and M_2 (d). The right side (e) presents the dice scores obtained by each method as a function of the number of samples in the database. Legend: blue - M_1, red - M_2.

These errors mainly stem from the small size of the database that weakens the power of the test. Instead, M_2 obtains much better and more steady scores, which demonstrate its robustness. M_2 performs better as we are able from a small database to increase the number of samples used for the comparison.

3.2   Experiments on Multiple Sclerosis Data

We have utilized the LONI ICBM database of healthy control diffusion images[1]. This database is composed of 160 control images: T1-weighted images (isotropic $1mm^3$) and diffusion images acquired on a 3T MRI scanner (b-value of 1000 s/mm^2, 30 directions with a resolution of 2x2x2 mm^3). This control subject database was compared to a database of 10 MS patient images acquired following a similar protocol with the same parameters. As a first step before processing, the diffusion tensor images are first registered to the T1-w images using a global affine transform [13] and a non linear free-form deformation [14] with few control points to recover EPI distortions. Then, a DTI atlas is computed from the control subjects DTI using Guimond's et al. atlas construction method [7], combined to a non linear tensor-based registration algorithm.

Each DTI patient image is then registered onto the atlas and compared voxel by voxel to the database of controls either with the method proposed in [5] M_1 or the proposed method M_2. We present in Fig. 2 a representative qualitative comparison of the results obtained by the two methods utilizing only a subgroup of 40 images from the controls subjects database.

We can notice on this figure that M_1 is affected by the small size of the database and the registration errors, resulting in a large number of false positive detections in Fig. 2.b. On the contrary, adding additional patches as it is done in M_2 leads to many more patches being considered (see Fig. 2.d) and possibly more accurate ones if the registration errors were in the bounds of the local neighborhood. As a consequence, the detection results in Fig. 2.c reveal much less false positives while keeping the detection power on the MS lesions.

Finally, we present a quantitative evaluation of the reproducibility of the obtained score maps when the control subjects database changes. To do so, we have, for each patient, repeatedly selected N_{Db} images out of the 160 images of the database. We have then computed for each of these sub-databases a score map deriving either from M_1 or M_2. To evaluate the variability of the scores, we have chosen to utilize the average of the voxelwise standard deviation of these maps. We present in Table 1 the average over all images of these standard deviation values for $N_{Db} = 20$, 40 and 80 images.

This table shows that the obtained standard deviations are significantly lower for M_2 (paired t-tests, p-value of 0.001). This indicates a better reproducibility of the results when considering our non-local approach. This confirms the robustness of the proposed method and the interest of utilizing neighboring patches, especially when performing a comparison against a very small database.

---

[1] https://ida.loni.ucla.edu/login.jsp?project=ICBM
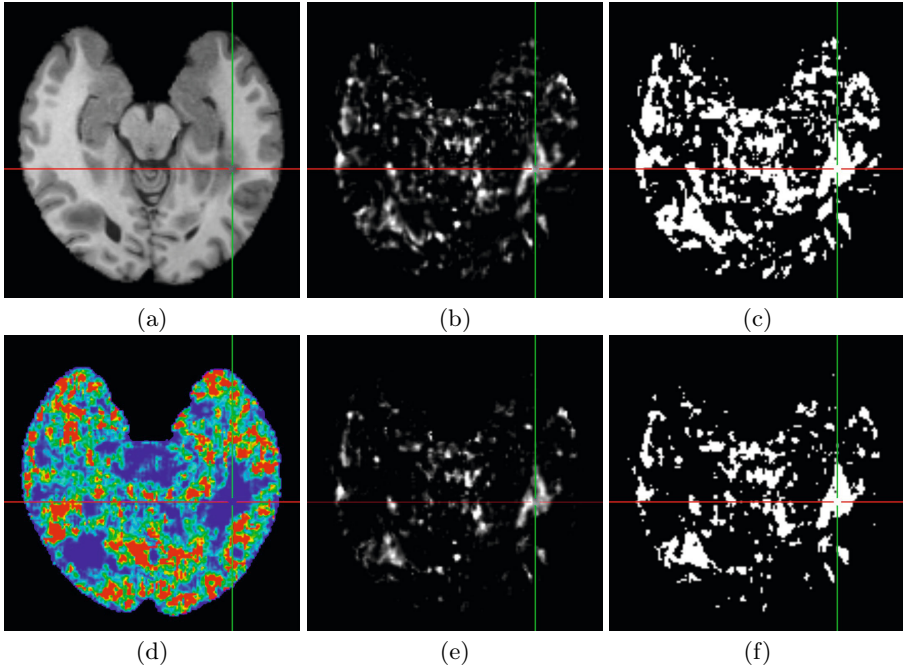
Fig. 2. Qualitative Comparison on Real MS Patient Images. Comparison of the score maps (Eq. (3)) and differences detected by the two methods M_1 (b,c) and M_2 (d-f). (a): T1 image of a patient, (b,e): score maps for M_1 and M_2, (d): number of patches kept for each voxel by M_2 (from blue: low number, to red: large number), (c,f): differences detected at the 95% level.

Table 1. Reproducibility of Comparison Results with Changing Databases. Average variation of z-scores over all voxels of all images for the compared methods.

	$N_{Db} = 20$	$N_{Db} = 40$	$N_{Db} = 80$
Method M_1	0.737	0.373	0.164
Method M_2	0.627	0.336	0.155

4   Conclusion

We have presented a new method for the robust detection of differences between a patient diffusion image and a population of control subject diffusion images. It relies on the search for additional patches in a local neighborhood of each voxel utilizing the non-local means framework adapted to diffusion tensor images in the Log-Euclidean space. We have demonstrated both on simulated and real datasets that this allows to detect more accurately differences even if the reference database is small, and to be more robust to potential registration errors. Moreover, it may be applied to any type of diffusion data that can be

represented as vector values such as ODFs in a spherical harmonics basis, which should further increase detection performance in regions with crossing fibers.

Future works will include an in-depth study of weights definition for oriented structures. The weights may be erroneous in patches where different orientations are present, which could lead to decreased performance. Accounting for these changes in orientations will therefore further improve comparison quality. We will also investigate other approaches to use the selected patches to detect differences. For example, our method could be coupled with a robust comparison algorithm such as the one proposed by Commowick et al. for tensors [15]. Accounting for spatial correlation between the selected patches could also bring further improvements to the comparison. Finally, we will also investigate how to extend our approach to robust population comparison.

References

1. Rovaris, M., Gass, A., Bammer, R., Hickman, S.J., Ciccarelli, O., Miller, D., Filippi, M.: Diffusion MRI in multiple sclerosis. Neurology 65(10), 1526–1532 (2005)
2. Filippi, M., Cercignani, M., Inglese, M., Comi, M.H.G.: Diffusion tensor magnetic resonance imaging in multiple sclerosis. Neurology 56, 304–311 (2001)
3. Lepore, N., Brun, C.A., Chou, Y.Y., Chiang, M.C., et al.: Generalized tensor-based morphometry of HIV/AIDS using multivariate statistics on deformation tensors. IEEE Transactions on Medical Imaging 27(1), 129–141 (2008)
4. Whitcher, B., Wisco, J.J., Hadjikhani, N., Tuch, D.S.: Statistical group comparison of diffusion tensors via multivariate hypothesis testing. Magnetic Resonance in Medicine (57), 1065–1074 (2007)
5. Commowick, O., Fillard, P., Clatz, O., Warfield, S.K.: Detection of DTI White Matter Abnormalities in Multiple Sclerosis Patients. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 975–982. Springer, Heidelberg (2008)
6. Goh, A., Lenglet, C., Thompson, P., Vidal, R.: A nonparametric riemannian framework for processing high angular resolution diffusion images and its applications to ODF-based morphometry. Neuroimage 56, 1181–1201 (2011)
7. Guimond, A., Meunier, J., Thirion, J.P.: Average brain models: A convergence study. Computer Vision and Image Understanding 77(2), 192–210 (2000)
8. Coupé, P., Yger, P., Prima, S., Hellier, P., Kervrann, C., Barillot, C.: An optimized blockwise non-local means denoising filter for 3D magnetic resonance images. IEEE Transactions on Medical Imaging 27(4), 325–441 (2008)
9. 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)
10. Buades, A., Coll, B., Morel, J.M.: A review of image denoising, with a new one. Multiscale Model. Simul. 4(2), 490–530 (2005)
11. Arsigny, V., Fillard, P., Pennec, X., Ayache, N.: Log-Euclidean metrics for fast and simple calculus on diffusion tensors. Magnetic Resonance in Medicine 56(2), 411–421 (2006)

12. Anderson, T.: An introduction to multivariate statistical analysis. Wiley (2003)
13. 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)
14. Rueckert, D., Sonoda, L.L., Hayes, C., Hill, D.L.G., Leach, M.O., Hawkes, D.J.: Nonrigid registration using free-form deformations: Application to breast MR images. IEEE Transactions on Medical Imaging 18(8), 712–721 (1999)
15. Commowick, O., Warfield, S.K.: A continuous STAPLE for scalar, vector and tensor images: An application to DTI analysis. IEEE Transactions on Medical Imaging 28(6), 838–846 (2009)

Group-Wise Consistent Fiber Clustering Based on Multimodal Connectional and Functional Profiles

Bao Ge[1], Lei Guo[1], Tuo Zhang[1], Dajiang Zhu[2], Kaiming Li[1], Xintao Hu[1], Junwei Han[1], and Tianming Liu[2]

[1] School of Automation, Northwestern Polytechnical University, Xi'an, China
{oct.bob,guolei.npu,zhangtuo.npu,dajiang.zhu,likaiming,
xintao.hu,junweihan2010}@gmail.com
[2] Department of Computer Science, University of Georgia, Athens, GA
tliu@cs.uga.edu

Abstract. Fiber clustering is an essential step towards brain connectivity modeling and tract-based analysis of white matter integrity via diffusion tensor imaging (DTI) in many clinical neuroscience applications. A variety of methods have been developed to cluster fibers based on various types of features such as geometry, anatomy, connection, or function. However, identification of group-wise consistent fiber bundles that are harmonious across multi-modalities is rarely explored yet. This paper proposes a novel hybrid two-stage approach that incorporates connectional and functional features, and identifies group-wise consistent fiber bundles across subjects. In the first stage, based on our recently developed 358 dense and consistent cortical landmarks, we identified consistent backbone bundles with representative fibers. In the second stage, other remaining fibers are then classified into the existing backbone bundles using their correlations of resting state fMRI signals at the two ends of fibers. Our experimental results show that the proposed methods can achieve group-wise consistent fiber bundles with similar shapes and anatomic profiles, as well as strong functional coherences.

Keywords: DTI, Fiber clustering, Resting state fMRI, Fiber Classification.

1 Introduction

Automatic fiber clustering based on diffusion tensor imaging (DTI) has become a very active research area for the purpose of group-based statistical analysis on the fiber bundles [3-9]. However, DTI tractography algorithms typically generate a large number (10,000–100,000) of fibers per subject, which makes the information provided by the fibers not easily comprehensible. Therefore, the large number of fibers is often grouped into fiber bundles by fiber clustering methods [3-9] to facilitate group-wise tract-based analysis.

A typical framework is to first define a similarity between pairwise fibers and then input the similarity matrix to standard data clustering algorithms. Therefore, various similarity measures have been proposed in the literature including geometric, anatomical,

connectional and functional characteristics of fibers [3-9]. For example, Maddah et al. [4] represented fibers as 3D quintic B-splines, Brun et al. [5] tried to capture the three geometric features using a 9-D descriptor, Corouge et al. [6] and Gerig et al. [7] proposed a mean closest distance that contains position and shape information, and Maddah et al. [3] enhances the Hausdorff similarity with Mahalanobis distance between fiber points. Later, anatomical (or atlas-based) feature was used to guide fiber clustering automatically/semi-automatically [11]. Recently, Ge et al. [8] made attempts to use the functional correlation derived from resting state fMRI (R-fMRI) data to guide fiber clustering.

This paper presents a novel two-stage hybrid fiber clustering approach that clusters fibers in a hierarchical way based on connectional and functional features. Specifically, the first stage groups a portion of fibers into group-wise consistent representative backbone bundles based on our recently developed 358 dense and consistent cortical landmarks [2]. The second stage classifies other remaining fibers into the backbone bundles obtained in the first stage according to functional coherences derived from R-fMRI data. The major advantage of this methodology is that those consistent and common 358 cortical landmarks define and form the reliable and corresponding backbone fiber bundles, which serve as the reliable reference for the following clustering of less consistent fibers. Furthermore, in the second stage, the traditional fiber clustering problem is converted into a fiber classification problem in which the functional coherence derived from R-fMRI data guides the fiber clustering procedure. In short, the proposed two-stage fiber clustering methodology effectively utilizes the deep-rooted common connectional and functional brain architectures to guide the fiber clustering processes such that the obtained fiber clusters possess both structural and functional correspondences across individuals, which was demonstrated by our experimental results.
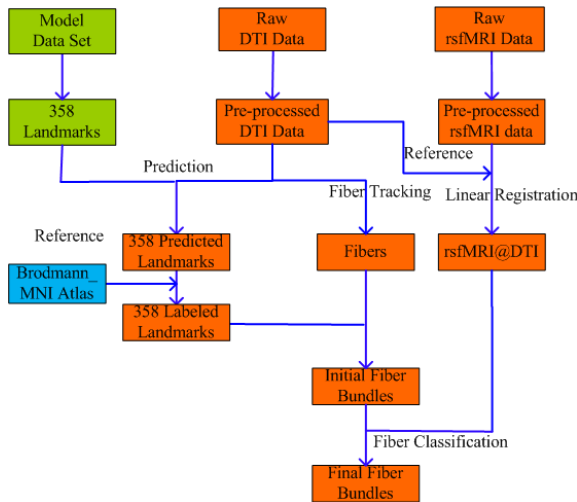
Fig. 1. Flowchart of the proposed computational framework

2   Materials and Methods

2.1   Overview

As summarized in Fig.1, our algorithmic pipeline includes the following steps. First, we pre-processed the raw DTI data and R-fMRI data, and then performed fiber tracking based on DTI data. Also, we registered the R-fMRI signals to the DTI space using FSL FLIRT. In the meantime, we predicted the 358 consistent landmarks for all subjects via the methods in [2], and grouped/labeled these 358 landmarks by the MNI (Montreal Neurological Institute) atlas. Then, we identified the backbone fiber bundles based on these consistent cortical landmarks, and each backbone bundle is represented by several representative fibers. Finally, we represented the backbone fiber bundles by the mean R-fMRI signals of these fibers, and classified other remaining fibers into these backbone bundles by comparing the wavelet-derived correlations of the R-fMRI signals.

2.2   Multimodal Data Acquisition and Pre-processing

Eight student volunteers were scanned using a 3T GE Signa MRI system under IRB approvals. We acquired the R-fMRI data with the dimensionality of 128*128*60*100, space resolution 2mm*2mm*2mm, TR 5s, TE 25ms, and flip angle 90 degrees. DTI data was acquired using the same spatial resolution as the R-fMRI data; the parameters were TR 15.5s and TE 89.5ms, with 30 DWI gradient directions and 3 B0 volumes acquired. Pre-processing of the R-fMRI data included brain skull removal, motion correction, spatial smoothing, temporal pre-whitening, slice time correction, global drift removal, and band pass filtering (0.01Hz~0.1Hz). For the DTI data, pre-processing steps included brain skull removal, motion correction, and eddy current correction. After the pre-processing, fiber tracking was performed using MEDINRIA (FA threshold: 0.2; minimum fiber length: 20). The DTI image space was used as the standard space from which to generate the tissue segmentation map and from which to show the functionally coherent fiber bundles on the cortical surface. DTI and fMRI images were registered via FSL FLIRT.

2.3   Identifying Backbone Fiber Bundles via 358 Consistent Cortical Landmarks

Recently, we identified and validated 358 group-wise consistent cortical landmarks that possess intrinsic correspondences across individuals and populations [2]. These landmarks have consistent DTI-derived fiber connection patterns and exhibit corresponding functional locations. Importantly, they have been reproduced in over 240 individual brains [2]. Thus, these 358 landmarks offer a universal and individuated brain reference system. In particular, these 358 landmarks can be accurately predicted in each individual brain with DTI data [2]. Figure 2(a) shows an example of these 358 landmarks.

Based on the MNI atlas, first, we grouped these landmarks into Brodmann-labeled classes, 37 Brodmann areas were used to label the 358 landmarks. We performed this step on one randomly chosen subject once, and then these labeled landmarks can be

applicable to other subjects because the 358 cortical landmarks possess correspondences. Figure 2(b) shows the MNI-labeled landmarks with different colors. Then, the backbone fiber bundles were identified base on the landmarks' labels that the fiber's two ends connect. That is, we grouped all of fibers connecting the same two Brodmann labels into the same backbone fiber bundle, thus we found the 32 common backbone bundles across 8 subjects by finding the common Brodmann labels across subjects. Quantitatively, these fibers count for approximately 6%~12% of all fibers in the whole brains of different subjects. Fig. 2(c) shows an example of backbone bundles.

(a)   (b)   (c)
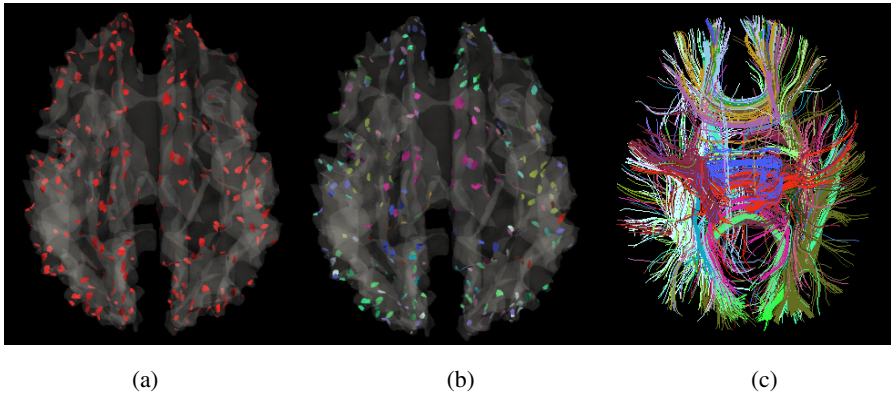
Fig. 2. (a) 358 landmarks (in red) from one randomly chosen subject. (b) The grouped/labeled landmarks via the MNI atlas. Each color represents one Brodmann area. (c) The grouped backbone fiber bundles. Each color represents one backbone bundle. The colors have no correspondences with those in (b).

2.4   Fiber Classification via Functional Coherence

The above identified backbone fiber bundles possess intrinsic correspondences across different brains because of the existing correspondences established by the 358 cortical landmarks [2]. Then, they are used as the common and reliable reference for the second stage of fiber clustering through fiber classification. The basic idea is that the remaining fibers are classified to one of the backbone bundles based on the Wavelet-based functional coherence derived from R-fMRI data.

Wavelet analysis is particularly well suited for the analysis of cortical fMRI time series in the resting state [13]. Specifically, the maximal overlap discrete wavelet transform (MODWT) is a redundant transform that is translation invariant and easy to compute using the pyramid algorithm. In particular, it does not suffer from the DWT's sensitivity to the choice of a starting point for a time series, and thus is adopted in this paper.

The steps of using MODWT for fiber classification are described as follows.

1. First, we extracted the R-fMRI signal from gray matter voxels that the fiber's two end points connect, using the method similar to [8]. One fiber has two time series signals, which were denoted by X_{i1}, X_{i2}, and the average value is X_i.

2. Then, the mean R-fMRI signal (X_{jmean}) within each backbone fiber bundle was computed. That is, $X_{jmean} = mean(X_i)$ for the *jth* bundle.
3. The MODWT was used to decompose each mean fMRI time series into the following scales or frequency intervals [10]: scale 1, 0.16–0.31 Hz; scale 2, 0.08–0.16 Hz; scale 3, 0.04–0.08 Hz; and scale 4, 0.02–0.04 Hz. Afterwards, we computed the MODWT wavelet coefficients ($\widetilde{W}_{j=3}^X$) on scale 3 of the wavelet decomposition.
4. For each other remaining fiber that is neighboring to the fibers within the backbone bundle, we computed the MODWT wavelet coefficients ($\widetilde{W}_{j=3}^Y$) in the same manner. Also, the wavelet correlation [1] between the fiber and each backbone fiber bundle was then computed as $|\rho_{XY}|$, and the fiber was classified into the backbone fiber bundle with maximal correlation value, which must be larger than a predefined threshold in order to ensure that the most relevant fibers are selected.

Notably, in this work, we focused on the scale 3 of the wavelet decomposition in that this is the frequency band most commonly studied in R-fMRI analyses and represents a reasonable trade-off between avoiding the physiological noise associated with higher frequency oscillations and the measurement error associated with estimating very low frequency correlations from limited time series [10]. And the threshold was chosen empirically. We manually selected the 11 fiber bundles according to the method in [12], and computed the functional correlation values between fibers within each bundle, then averaged them for all 11 fiber bundles. The averaged value of 0.7 was chosen as threshold.

3   Experimental Results

In total, we identified 32 group-wise consistent fiber bundles for the whole brain, as shown in Fig 3. For the purpose of visual differentiation, each fiber bundle was represented by the representative fiber (shown in Fig 3(b)) whose mean closest distance with other fibers within the bundle was minimal. Each corresponding fiber bundle and the representative fiber in Fig. 3 have the same color in different brains. It is evident that the distributions of these 32 representative fibers are quite reasonable and consistent. As a more detailed example, Fig.4 shows 8 consistent fiber bundles from two randomly selected subjects. For a quantitative comparison, we computed the Hausdorff distances between the corresponding representative fibers of the eight subjects, as shown in Table 1. It can be seen that the Hausdorff distances are relatively small.

Moreover, we compared the percentages of streamline fibers in the 8 backbone fiber bundles and those of the corresponding finally clustered bundles after the second stage classification in Table 2. On average, the percentage of consistently clustered fibers increases from 8.03% to 25.78%, suggesting that the group-wise consistent backbone fiber bundles can really serve as the common white matter fiber tracts for clustering, and not all fibers were clustered into these 32 bundles because the 358 landmarks can only cover a portion of the cortex. Table 3 shows the functional coherence of the 8 final fiber bundles with the corresponding backbone fiber bundles. We can see that final fiber bundles maintain the high functional coherence after classification. These above results demonstrated the fiber bundles have both similar connection patterns and functional coherences.

(a)

(b)
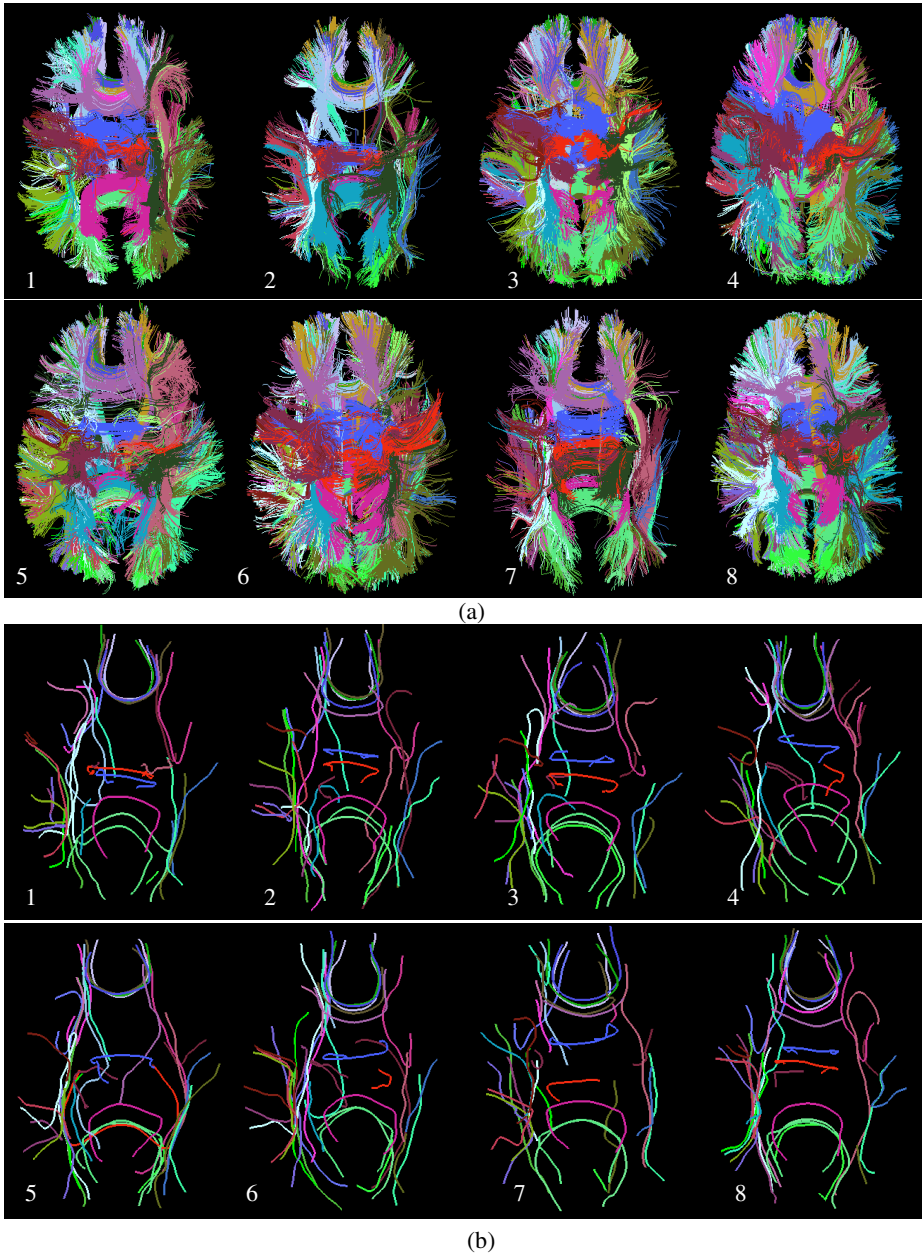
Fig. 3. The 32 group-wise consistent fiber bundles for 8 subjects. (a) The 32 fiber bundles with the corresponding colors across subjects. (b) The 32 representative fibers of bundles for each subject with the same corresponding colors as in (a).
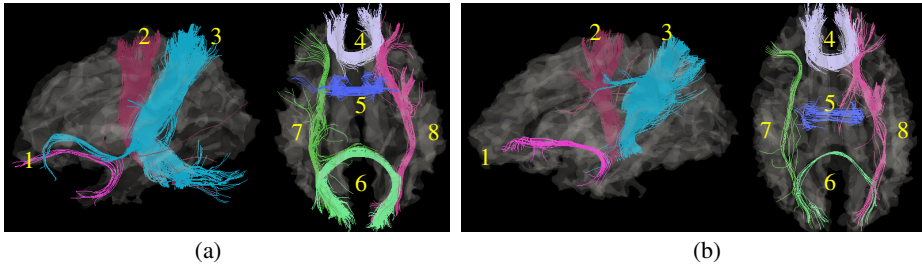
Fig. 4. The detailed visualization of 8 consistent fiber bundles from 2 subjects (a) and (b). The same corresponding fiber bundle (indexed by numbers) is in the same color.

Table 1. Hausdorff distances (mm) between the 8 consistent corresponding fiber bundles across 8 subjects

	$Dis_{s1,s2}$	$Dis_{s1,s3}$	$Dis_{s1,s4}$	$Dis_{s1,s5}$	$Dis_{s1,s6}$	$Dis_{s1,s7}$	$Dis_{s1,s8}$
1	2.790	4.648	3.121	5.070	4.035	3.459	2.161
2	3.347	4.668	6.256	4.655	4.192	4.068	5.264
3	5.877	4.934	4.226	5.637	6.300	3.822	4.804
4	3.301	5.264	3.041	3.827	2.556	2.747	2.443
5	3.469	3.386	4.191	2.590	4.026	1.903	4.330
6	2.193	2.616	4.698	2.807	2.397	2.398	5.395
7	6.951	5.178	6.874	5.326	6.091	3.268	6.192
8	5.289	2.804	6.938	5.666	7.673	6.708	6.057

Table 2. The mean percentages of the 8 backbone fiber bundles (row #2) among all fibers in the whole brain, and those of the finally clustered fiber bundles (row #3). The total number of all the backbone fibers   for all subjects is also shown (column #10).

	1	2	3	4	5	6	7	8	all
Initial	0.18%	0.72%	0.57%	0.18%	0.22%	0.14%	0.08%	0.10%	8.03%
Final	0.22%	2.06%	1.90%	1.01%	0.92%	0.35%	0.27%	0.30%	25.78%

Table 3. Functional coherence of the 8 final fiber bundles. The functional coherence was computed as the mean correlation of each fiber's R-fMRI signal (X_i) within the final fiber bundles and the mean R-fMRI signal (X_{jmean}) within each backbone fiber bundle.

	1	2	3	4	5	6	7	8
coherence	0.8065	0.7283	0.7538	0.9038	0.8347	0.8034	0.7573	0.7303

4   Conclusion

This paper proposed a new fiber clustering method that achieves group-wise consistent fiber bundles in two stages. Conceptually, this methodology utilizes the common structural and functional brain architectures inferred from DTI and R-fMRI data. In particular, the 358 cortical landmarks provide a basis for the extraction of backbone

fiber bundles, whose correspondences and consistencies are achieved automatically. In the second stage, the functional coherences derived from R-fMRI data were used to guide the classification of the remaining fibers into the already consistent backbone bundles. Both qualitative and quantitative analyses demonstrated the good performance of the proposed framework. In the future, we plan to investigate finer scale clustering of these fibers into a more structurally and functionally homogenous bundles. In addition, we plan to perform large scale task-based fMRI studies to validate the functional correspondences of these backbone fiber bundles.

References

1. Percival, D.B., Walden, A.T.: Wavelet methods for time series analysis. Cambridge UP, Cambridge (2000)
2. Zhu, D., et al.: DICCCOL: Dense Individualized and Common Connectivity-based Cortical Landmarks. Cerebral Cortex (accepted, 2012)
3. Maddah, M., Grimson, W., Warfield, S., Wells, W.: A unified framework for clustering and quantitative analysis of white matter fiber tracts. Med. Image Anal. 12(2), 191–202 (2008)
4. Maddah, M., Grimson, W., Warfield, S.: Statistical Modeling and EM Clustering of White Matter Fiber Tracts. In: ISBI, vol. 1, pp. 53–56 (2006)
5. Brun, A., Knutsson, H., Park, H.-J., Shenton, M.E., Westin, C.-F.: Clustering Fiber Traces Using Normalized Cuts. In: Barillot, C., Haynor, D.R., Hellier, P. (eds.) MICCAI 2004. LNCS, vol. 3216, pp. 368–375. Springer, Heidelberg (2004)
6. Corouge, I., Gouttard, S., Gerig, G.: Towards a Shape Model of White Matter Fiber Bundles Using Diffusion Tensor MRI. In: ISBI, pp. 344–347 (2004)
7. Gerig, G., Gouttard, S., Corouge, I.: Analysis of Brain White Matter via Fiber Tract Modeling. In: IEEE EMBS, vol. 2, pp. 4421–4424 (2004)
8. Ge, B., Guo, L., Lv, J., Hu, X., Han, J., Zhang, T., Liu, T.: Resting State fMRI-Guided Fiber Clustering. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part II. LNCS, vol. 6892, pp. 149–156. Springer, Heidelberg (2011)
9. Ge, B., Guo, L., Li, K., Li, H., Faraco, C., Zhao, Q., Miller, S., Liu, T.: Automatic Clustering of White Matter Fibers via Symbolic Sequence Analysis. In: SPIE Medical Image, vol. 7623, pp. 762327.1–762327.8 (2010)
10. Fornito, A., Zalesky, A., Bullmore, E.: Network Scaling Effects in Graph Analytic Studies of Human Resting-State fMRI Data. Front. Syst. Neurosci. 4, 22 (2010)
11. O'Donnell, L.J., Westin, C.-F.: Automatic tractography segmentation using a high-dimensional white matter atlas. IEEE Transactions on Medical Imaging 26(11), 1562–1575 (2007)
12. Wakana, S., et al.: Reproducibility of quantitative tractography methods applied to cerebral white matter. NeuroImage 36(3), 630–644 (2007)
13. Maxim, V., Sendur, L., Fadili, M.J., Suckling, J., Gould, R., Howard, R., Bullmore, E.T.: Fractional Gaussian noise, functional MRI and Alzheimer's disease. NeuroImage 25, 141–158 (2005)

Learning a Reliable Estimate of the Number of Fiber Directions in Diffusion MRI

Thomas Schultz

Max Planck Institute for Intelligent Systems, Tübingen, Germany⋆

Abstract. Having to determine an adequate number of fiber directions is a fundamental limitation of multi-compartment models in diffusion MRI. This paper proposes a novel strategy to approach this problem, based on simulating data that closely follows the characteristics of the measured data. This provides the ground truth required to determine the number of directions that optimizes a formal measure of accuracy, while allowing us to transfer the result to real data by support vector regression. The method is shown to result in plausible and reproducible decisions on three repeated scans of the same subject. When combined with the ball-and-stick model, it produces directional estimates comparable to constrained spherical deconvolution, but with significantly smaller variance between re-scans, and at a reduced computational cost.

1 Introduction

Multi-compartment models are a traditional way of estimating more than a single fiber orientation in diffusion MRI [1,2]. The number of fiber compartments used in such models can have a profound effect on the estimated directions, making it mandatory to decide on a setting that is adequate for any given voxel.

Despite this, only few systematic approaches to this problem are available: Automated Relevance Determination [2] has been used to force the weights of fiber compartments with insufficient statistical support to zero. However, it requires computation in a full Bayesian framework. The Bayesian Information Criterion has been demonstrated to produce suboptimal results even on idealized synthetic data [3]. Approaches based on peaks of the fiber ODF [3,4] require setting a threshold, and suffer from noise-induced spurious peaks.

This work proposes a novel strategy for setting the number of fiber compartments. It formalizes the pragmatic view that it is best to use the number that leads to the most accurate estimates of the desired parameters. Since the in vivo data lacks ground truth, the optimal number of compartments is determined in simulated data that closely follows the characteristics of the experimental data. Based on this synthetic data, a classifier is trained to recognize voxels that are best analyzed by a ball-and-stick model [2] with a single, two, or three fiber compartments, and thereby achieves reliable and efficient fiber direction estimates.

2   Defining the "Most Adequate" Number of Directions

When analyzing a diffusion MRI signal that arises as an average over n fiber compartments with principal directions $\mathbf{v}_i$ and relative weights w_i, two or more of the $\mathbf{v}_i$ may be so close to collinearity that, given noisy measurements with limited angular and spectral resolution, it becomes impossible to separate them with reasonable precision. In this case, we prefer to describe them with a single estimate $\hat{\mathbf{v}}_j$. Similarly, when some of the weights w_i are so small that the associated directions can no longer be reliably estimated, we prefer to set them to zero in order to obtain a less complex model that is more robust to noise.

This tradeoff is formalized by the following definition of weighted average angular error (WAAE), which measures the average angle between each true direction $\mathbf{v}_i$ and its nearest estimate $\hat{\mathbf{v}}_j$, weighted by its true volume fraction w_i:

$$\text{WAAE} := \sum_{i=1}^{n} w_i \min_{j} \arccos(|\mathbf{v}_i \cdot \hat{\mathbf{v}}_j|) \tag{1}$$

Given a model that requires choosing the number of fiber compartments, we define the number that minimizes WAAE as the one most suitable for analysis. Even though WAAE is used throughout this paper, the fundamental idea is to use some formal error measure like it to decide on the "right" number of directions, not necessarily its exact definition. In particular, when dealing with advanced multi-compartment models that additionally estimate parameters such as axon diameter [5] or account for fiber fanning or bending, learning could be based on an objective function that penalizes errors in those.

3   Learning the Number of Fiber Compartments

Since evaluation of WAAE requires ground truth, it cannot be computed directly for in vivo data. Instead, we use machine learning to train a classifier that predicts the best number of fiber compartments for experimental data based on its similarity to simulated data, for which WAAE can be computed.

3.1   Support Vector Regression

Given a suitable representation $\mathbf{x}$ of the diffusion-weighted signal $S(\theta, \phi)$, we seek a function $\hat{f}(\mathbf{x})$ whose value provides an estimate of the most adequate number of directions to analyze S. We obtain $\hat{f}(\mathbf{x})$ through support vector regression, which produces functions of the form

$$\hat{f}(\mathbf{x}) = \sum_{i=1}^{\ell} (\alpha_i - \alpha_i^*) k(\mathbf{x}_i, \mathbf{x}) + b. \tag{2}$$

Support vector regression requires us to decide on a representation $\mathbf{x}$ of the signal, to specify training samples $\mathbf{x}_j$ for which the value of $f(\mathbf{x}_j)$ is known,

and to specify a kernel function $k(\mathbf{x}_j, \mathbf{x})$ that measures the similarity between any input $\mathbf{x}$ and the training data $\mathbf{x}_j$. Based on this, an optimization procedure automatically finds a subset $\{\mathbf{x}_i\}_{i=1}^{\ell}$ of the training data and suitable parameters α_i, α_i^*, and b, so that a given distance measure between $\hat{f}(\mathbf{x}_j)$ and the training values $f(\mathbf{x}_j)$ is minimized, subject to additional constraints. Full details of the method are beyond the scope of this paper and are given in [6].

3.2   Feature Definition and Kernel Selection

The representation $\mathbf{x}$ from which Eq. (2) is computed is known as a feature vector. Ideally, it should encode the available prior knowledge about f. For example, the number of fiber compartments clearly should not be affected by joint rotations of all fibers. Therefore, $\mathbf{x}$ should be invariant under rotations of the measurement frame, while providing information that can be used to infer a suitable number of directions.

The experiments reported in this paper use the three sorted eigenvalues of a single diffusion tensor per voxel as the feature vector. Even though the diffusion tensor is insufficient to resolve the *directions* of more than a single compartment, the results show that it reliably indicates their *number.* Experiments with more complex features based on rotational invariants of various HARDI models, including ADC profiles [7], Q-Balls, and fiber ODFs, have led to small improvements on simulated data, but reproducibility on real data was much reduced.

As the kernel $k(\mathbf{x}_j, \mathbf{x})$, we use a standard radial basis function, $k(\mathbf{x}_j, \mathbf{x}) = \exp(-\gamma \|\mathbf{x}_j - \mathbf{x}\|^2)$. The parameter γ, as well as two additional parameters, C and ν, which control the number ℓ of support vectors in Eq. (2), have been fixed automatically using cross-validation [6].

3.3   Training Data and Labels

In order to successfully transfer the learned function $\hat{f}(\mathbf{x})$ to real data, it is essential that the characteristics of the training data $\mathbf{x}_j$ be as similar as possible to the experimental data. Therefore, we generate it directly from the measurements, with random weights $w_i > 0$, $w_1 + w_2 + w_3 = 1$, and random directions $\mathbf{v}_i$. Note that even though we always simulate three compartments, modeling them with only one or two compartments may lead to a smaller WAAE when some w_i are very small, or at least two $\mathbf{v}_i$ are nearly collinear.

As in [8], the N voxels with largest FA in the measured data are assumed to contain a single dominant direction, indicated by the principal diffusion direction. It is our goal to reproduce the natural variability of these single fiber voxels, including the bending and spreading which is present even in the most anisotropic voxels. Therefore, we randomly select one of the N voxels as a template for each simulated fiber compartment. In order to get a sufficiently realistic estimate of the variability, we use $N = 1000$, corresponding to $\mathrm{FA}_{\min} \approx 0.68$.

Since the classifier should be able to handle partial voluming with cerebrospinal fluid (CSF) or gray matter, we include it in our training data. For this, we use the N *least* anisotropic voxels each from a CSF mask, obtained by Otsu

segmentation [9] on mean diffusivity, and a brain mask excluding CSF, respectively. Each simulated voxel $\mathcal{V}$ selects a near-isotropic signal $I(\theta, \phi)$ from either gray matter or CSF, and weights it by a uniform random factor $w_{\mathrm{iso}} \in [0, 0.5]$.

Taken together, the simulated signal S is given by the following weighted combination of a randomly selected near-isotropic signal I, three randomly selected single-fiber signals R_i, rotated to match the directions $\mathbf{v}_i$, and error terms ε that are estimated using residual bootstrapping [10]:

$$S(\theta, \phi) = w_{\mathrm{iso}}(I(\theta, \phi) + \varepsilon_{\mathrm{iso}}) + (1 - w_{\mathrm{iso}}) \sum_{i=1}^{3} w_i \left(R_i(\theta, \phi) + \varepsilon_i \right) \qquad (3)$$

From the simulated data, eigenvalue features $\mathbf{x}_j$ are computed, and directional estimates $\hat{\mathbf{v}}_j$ are obtained from the ball-and-stick model [2] with one, two, and three fiber directions. The number of compartments that led to the smallest WAAE is used as the true value of $f(\mathbf{x}_j)$ in the support vector regression. No cases other than $\{1, 2, 3\}$ are considered, since we assume a separate classifier defines a white matter mask, and no prior work has plausibly reconstructed more than three fiber directions from a single voxel.

Since we only use discrete values $f(\mathbf{x}_j) \in \{1, 2, 3\}$ in our training data, we should avoid including boundary cases in which two settings perform similarly well. Therefore, we simulate a large number of voxels (250,000) and only train on the 1,000 examples $\mathbf{x}_j$ of each class for which the difference in WAAE between the optimal choice of compartment number and the second-best choice was largest.

4   Results on Experimental Data

Three repeated diffusion MR acquisitions have been obtained within the same session at 3 T, with voxel size $2 \times 2 \times 2\,\mathrm{mm}^2$, 71 gradient directions, 8 B0 images, $b = 1000\,\mathrm{s/mm}^2$. Eddy current distortions and head motion have been corrected for using FSL (www.fmrib.ox.ac.uk), and the B matrix has been rotated accordingly [11]. In order to evaluate reproducibility, results that use all available data are compared to results obtained from the three individual repeats.

4.1   Number of Fibers

We clamp the values of $\hat{f}(\mathbf{x})$ to range $[1, 3]$ and round them to obtain discrete classes $\{1, 2, 3\}$. The percentages of voxels that were marked as being best analyzed with one, two, or three fiber compartments are listed in Table 1 for a brain mask (excluding CSF), and for two different thresholds of Fractional Anisotropy.

A comparison to constrained spherical deconvolution (with $l_{\max} = 8$) was performed using the software MRtrix [8]. Applying an FOD threshold of 0.1 as in [4] results in a much smaller number of single fiber voxels than with our classifier. Moreover, compared to the combined data from all three repeats, deconvolution systematically reported a smaller number of single-fiber voxels in the individual repeats (between 2.3% and 2.4%, rather than 3.4%).

Table 1. Counting the peaks in constrained deconvolution (CSD) estimates fewer single-fiber voxels than our classifier, consistent with known effects of noise on CSD

	Learned Classifier			Non-Neg. Deconvolution		
	non CSF	FA > 0.1	FA > 0.2	non CSF	FA > 0.1	FA > 0.2
1 Fiber	7.6 %	9.8 %	15.8 %	3.4 %	4.4 %	7.3 %
2 Fibers	20.3 %	26.1 %	43.6 %	18.6 %	24.1 %	40.6 %
3 Fibers	72.1 %	64.1 %	40.6 %	78.0 %	71.5 %	52.1 %

(a)   (b)   (c)
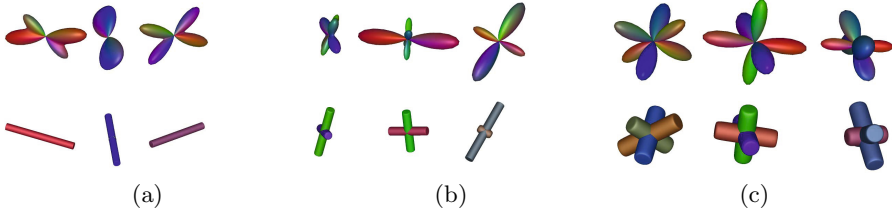
Fig. 1. While fiber ODFs from constrained deconvolution (top row) show reasonable agreement with the ground truth (bottom row) in simulated two- and three-fiber cases (b/c), large spurious peaks arise in two out of the three shown single directions (a)

This indicates that the reduced effective SNR in the individual measurements leads to spurious peaks in the fiber ODFs, which reduces the reliability of their number as an indicator of distinct fiber compartments. Figure 1 illustrates the problem using three examples each from the training sets that define the single-, two-, and three-fiber cases (a–c). The rods in the bottom row indicate the ground truth directions $\mathbf{v}_i$, lengths indicating relative weights w_i (radii being reduced in (a) and (b) to avoid occluding the smaller contributions in (b)).

The strong spurious peaks that arise in two out of the three single-fiber cases (a) are a known problem of deconvolution, particularly in the presence of partial voluming with isotropic compartments [12]. They are caused by the fact that the regularization introduced in [8] suppresses negative peaks, but does not address spurious positive peaks. In contrast to this, the fraction of fibers assigned to each class by our classifier was stable across re-scans. The percentage of single-fiber voxels in the individual repeats varied between 7.5% and 7.9%.

The reproducibility of our estimates is further confirmed by Table 2, which lists the percentage of voxels in which the classification in each individual measurement agreed with the class assigned based on the combined data, as well as the mean difference and the 95% confidence interval of the value of $\hat{f}(\mathbf{x})$.

Figure 2 presents a visual comparison between our classification (a/c) and the results from fODF thresholding (b/d) in a coronal (a/b) and an axial (c/d) slice. As in [4], red, green, and blue indicate one, two, and three directions, respectively. In (a/c), the values of $\hat{f}(\mathbf{x})$ are mapped before rounding, to demonstrate the smooth transition between the classes.

Table 2. Good agreement between classification of the combined data and individual repeats supports the reproducibility of our method

	Learned Classifier			Non-Neg. Deconvolution		
	Agreement	Mean	95 % Conf.	Agreement	Mean	95 % Conf.
1st Repeat	94.5 %	0.062	0.215	89.7 %	0.106	1.0
2nd Repeat	95.3 %	0.056	0.189	90.0 %	0.103	1.0
3rd Repeat	95.3 %	0.056	0.192	90.1 %	0.102	1.0
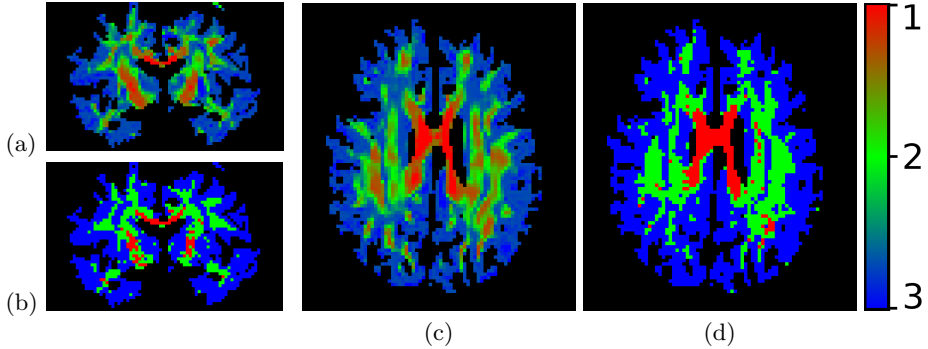
Fig. 2. Compared to the fiber number estimate from constrained deconvolution (b/d), our classifier (a/c) provides more coherent clusters of single fiber voxels (red), and smooth transitions between the classes

4.2   Estimates of Fiber Direction

Since the parameter we are ultimately interested in is fiber orientation, let us now consider the reproducibility of directional estimates based on the ball-and-stick model [2] when the number of sticks is determined by our function $\hat{f}(\mathbf{x})$.

Table 3 reports the weighted average angular deviation (WAAD) within a white matter mask (FA > 0.2). It is computed from Eq. (1), by treating the estimates from the combined data as "ground truth". The results from all three repeats were very similar and have been averaged for presentation. Compared to constrained deconvolution, ball-and-stick achieves slightly lower precision in the two-fiber case, but higher reproducibility in the one- and three-fiber cases. The unfavorable 90% confidence bound on deconvolution-based single fiber estimates is consistent with the emergence of spurious peaks as observed in Figure 1 (a).

The mean WAAD over the whole white matter was 9.8° when combining ball-and-stick with our classifier, which improves over the WAAD achieved by deconvolution (11.0°). According to a two-sided t-test on the distribution of WAADs from all white matter voxels, this difference is highly significant ($p < 10^{-20}$), in each of the three repeated measurements. In contrast, fitting the ball-and-stick model with three compartments in all voxels led to a larger average WAAD (11.9°). This confirms the importance of selecting an adequate number of fibers based on the data.

Table 3. Reasonable agreement is achieved between direction estimates from the combined data and individual measurements for one- and many two-fiber voxels

	Ball-and-Stick			Non-Neg. Deconvolution		
	Mean	Median	90 % Conf.	Mean	Median	90 % Conf.
1 Fiber	2.0	1.2	3.0	4.0	1.9	10.5
2 Fibers	8.7	5.7	20.9	7.1	5.4	14.7
3 Fibers	14.1	13.5	23.6	15.0	14.6	24.6

(a) Ball-and-Stick Result   (b) Constrained Deconvolution Result
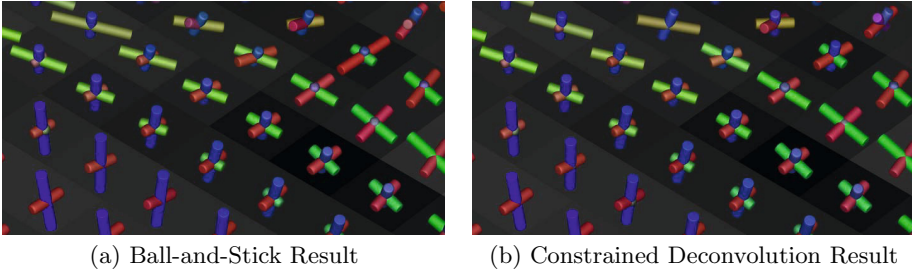
Fig. 3. The triple crossing between corpus callosum (red), corticospinal tract (blue) and superior longitudinal fasciculus (green) was the only region that produced reproducible three-fiber estimates. Closeup shows a slanted view onto an axial slice.

Large absolute errors indicate that many of the voxels that have been labeled as "three-fiber" by both methods do not afford reliable directional estimates using either model. However, inspecting all voxels in which three-fiber estimates were consistently obtained with less than 10° WAAD revealed two clear clusters: The triple crossing between corpus callosum, corticospinal tract, and superior longitudinal fasciculus, in both hemispheres. Visual inspection of a detail of that region in Figure 3 suggests that, when combined with an appropriate classifier, ball-and-stick fitting produces directional estimates that are very similar to constrained deconvolution.

The proposed method is computationally efficient. It took less than 10 seconds to propose a fiber number for all 96,000 voxels within the brain mask. Subsequent fitting of the ball-and-stick model took 74 seconds. In comparison, constrained deconvolution and subsequent peak finding with the implementation from [8] took almost 10 minutes on the same 2.7 GHz workstation.

5   Conclusion

A novel strategy to select the number of fiber directions in multi-compartment models has been presented, which explicitly aims to minimize a formal measure of error in the estimated model parameters. Since accuracy cannot be measured in vivo, a machine learning approach is used to automatically transfer insights from simulated data with known ground truth to experimental data.

On the conceptual side, the main contribution of this work is to demonstrate that such a transfer produces plausible and reproducible results. As a practical benefit, combining the proposed classifier with the ball-and-stick model produces directional estimates that are similar overall, but more reproducible than the ones from constrained deconvolution across re-scans, particularly in single-fiber voxels. They are also obtained at a markedly reduced computational cost. This seems particularly relevant when performing bootstrapping-based tractography, which requires repeated model fitting [13].

References

1. Tuch, D.S., Reese, T.G., Wiegell, M.R., Makris, N., Belliveau, J.W., Wedeen, V.J.: High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity. Magnetic Resonance in Medicine 48, 577–582 (2002)
2. Behrens, T.E.J., Johansen-Berg, H., Jbabdi, S., Rushworth, M.F.S., Woolrich, M.W.: Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? NeuroImage 34, 144–155 (2007)
3. Schultz, T., Westin, C.-F., Kindlmann, G.: Multi-Diffusion-Tensor Fitting via Spherical Deconvolution: A Unifying Framework. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part I. LNCS, vol. 6361, pp. 674–681. Springer, Heidelberg (2010)
4. Jeurissen, B., Leemans, A., Tournier, J.D., Jones, D.K., Sijbers, J.: Estimating the number of fiber orientations in diffusion MRI voxels: a spherical deconvolution study. In: Proc. Int'l Society of Magnetic Resonance in Medicine, ISMRM (2010)
5. Zhang, H., Dyrby, T.B., Alexander, D.C.: Axon Diameter Mapping in Crossing Fibers with Diffusion MRI. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part II. LNCS, vol. 6892, pp. 82–89. Springer, Heidelberg (2011)
6. Schölkopf, B., Smola, A.J.: Learning with Kernels. MIT Press (2002)
7. Alexander, D.C., Barker, G.J., Arridge, S.R.: Detection and modeling of non-gaussian apparent diffusion coefficient profiles in human brain data. Magnetic Resonance in Medicine 48, 331–340 (2002)
8. Tournier, J.D., Calamante, F., Connelly, A.: Robust determination of the fibre orientation distribution in diffusion MRI: Non-negativity constrained super-resolved spherical deconvolution. NeuroImage 35, 1459–1472 (2007)
9. Otsu, N.: A threshold selection method from gray-level histograms. IEEE Trans. on Systems, Man and Cybernetics 9(1), 62–66 (1979)
10. Chung, S., Lu, Y., Henry, R.G.: Comparison of bootstrap approaches for estimation of uncertainties of DTI parameters. NeuroImage 33(2), 531–541 (2006)
11. Leemans, A., Jones, D.K.: The B-matrix must be rotated when correcting for subject motion in DTI data. Magnetic Resonance in Medicine 61, 1336–1349 (2009)
12. Dell'Acqua, F., Scifo, P., Rizzo, G., Catani, M., Simmons, A., Scotti, G., Fazio, F.: A modified damped richardson-lucy algorithm to reduce isotropic background effects in spherical deconvolution. NeuroImage 49, 1446–1458 (2010)
13. Jeurissen, B., Leemans, A., Jones, D.K., Tournier, J.D., Sijbers, J.: Probabilistic fiber tracking using the residual bootstrap with constrained spherical deconvolution. Human Brain Mapping 32, 461–479 (2011)

Finding Similar 2D X-Ray Coronary Angiograms

Tanveer Syeda-Mahmood[1], Fei Wang[1], R. Kumar[1], D. Beymer[1], Y. Zhang[1],
Robert Lundstrom[2], and Edward McNulty[2]

[1] IBM Almaden Research Center, San Jose, CA, USA
[2] Kaiser San Francisco Medical Center, San Francisco, CA, USA

Abstract. In clinical practice, physicians often exploit previously observed patterns in coronary angiograms from similar patients to quickly assess the state of the disease in a current patient. These assessments involve visually observed features such as the distance of a junction from the root and the tortuosity of the arteries. In this paper, we show how these visual features can be automatically extracted from coronary artery images and used for finding similar coronary angiograms from a database. Testing on a large collection has shown the method finds clinically similar coronary angiograms from patients with similar clinical history.

1 Introduction

X-ray Coronary angiography is a commonly used technique to assess the state of coronary artery disease (CAD). During assessment, clinicians look for characteristic visual features, taking into account the overall disease burden, the complexity of individual lesions, and placing more weight on proximal stenoses of the coronary arteries. Even though there are quantitative assessment scores such as the Syntax Score[12], they require manual input of angiographic information. Thus the clinicians still characterize the disease by 'eyeballing' on salient visual features such as lumen variation or the relative thickness of arteries (see Fig. 1a-c)[5], the distance of the junctions from the root, the number of trifurcations, etc. In this paper, our goal is to automatically extract features from coronary angiograms that mimic this process and learn a distance matrix to retrieve similar coronary angiograms for purposes of clinical decision support.

Automatically deriving such salient features from coronary artery imaging is, however, a challenging problem. It requires reliable separation of the major coronary arteries from the background. Complete delineation of arteries is difficult due to the similarities in intensity distribution in the regions surrounding the arteries. It also requires a reliable detection of all major junctions and the tubular arterial segments between junctions to allow computation of features such as arterial width and curvature changes or tortuosities.

2 Related Work

Much of the existing work on coronary angiogram analysis has focused on the preprocessing and segmentation of coronary angiograms. The majority of the

<div align="center">(a)   (b)   (c)</div>
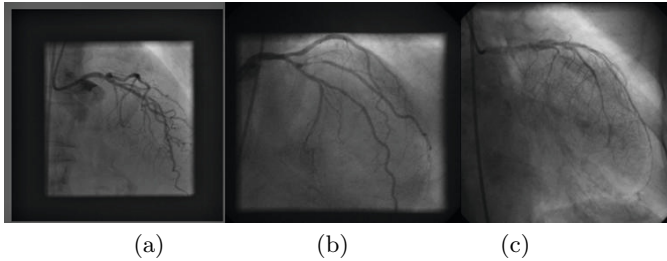
Fig. 1. Observable semantic features in coronary angiograms. (a) The dark blobs along arterial walls are calcifications, and irregular thickness variations can be observed.(b) Lumen variations along arteries. (c) Thinner/fine arteries showing diffused vessels.

methods locate vessel regions using various filters, deformable model methods, and supervised or unsupervised learning-based approaches [3,4,10,2]. Completely automatic artery tree extraction has also been attempted in 3D CT Angiography [15], but often rely on user identification of root in 2D X-ray angiography [7]. Previous work has also studied the junction detection in arteries. While the majority of the work is on 3D CT angiography data, relying on a good 3D vessel tree model and a robust 2D-3D shape alignment algorithm, [1,15], junction detection in 2D X-ray angiography has been restricted to either sensing Y or X junctions in pixel neighborhoods or using the intersection of artery centerlines.

There is also work on quantitative characterization of coronary artery disease in the identification of coronary artery root. Popular angiography tools offer measurements such as luminal cross-sectional area and percentage area stenosis. Most of these tools, however, require some manual assistance including the identification of the coronary tree root.

3   Image Pre-processing of Coronary Angiograms

Our pre-processing uses well-known techniques put together in a new sequence to delineate coronary arteries. Since the clinical assessment focuses on major coronary arteries, accurate and complete tree reconstruction is not necessary.

3.1   Highlighting Coronary Arteries

Starting from coronary angiogram video frames, we first extract a region of interest containing the arteries by exploiting the spatial and temporal variance in pixels. We then highlight the coronary artery vessel structure using a suitable ridge detection filter. While several filters could be used [4], we adopted Radon-Like Features (RLF)[6] as it does a non-isotropic sampling of neighborhoods based on edge sensing along different orientations. It has been shown to give a more complete highlighting of coronary arteries, including minor segments, while still suppressing noise. The result of RLF-filtering the coronary artery image of Fig. 2a is shown in Fig. 2c.
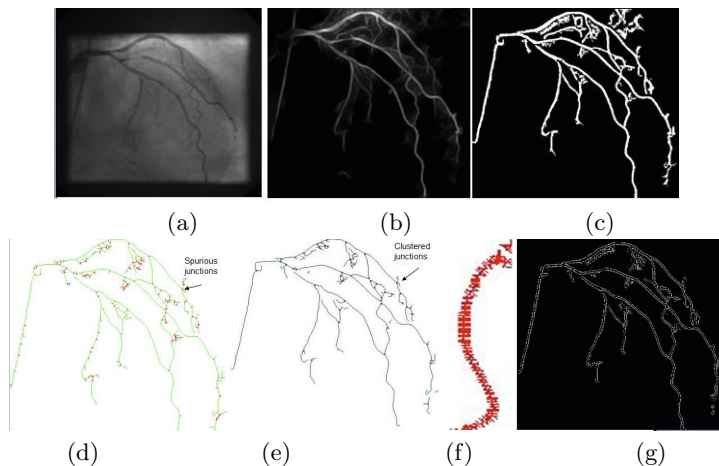
Fig. 2. Illustration of artery preprocessing.(a) original image. (b) RLF-filtered image. (c) Binary image obtained by adaptive thresholding. (d) Skeletonization and initial junction detection. (e) Junction clustering. (f) Tubular region detection for a single skeletal curve. (g)Tubular segments surrounding skeletal curves between junctions.

3.2 Locally Adaptive Statistical Thresholding

Next, we threshold the intensity gradients in the filtered region to separate the foreground vessel region from the background. Since a global threshold is insufficient, we model the filtered coronary image as the output of a short-space stationary process. Since coronary arteries have small thickness (less than 16 pixels), an overlapping window analysis with a small window size $W \times W$ ($W \le 16$), is sufficient. Within each window of size $W \times W$, we find an optimal threshold T using similar ideas to Otsu thresholding [8], such that it separates the pixels within the regions into two classes with minimized intra-class variance $\sigma_w^2(T) = \omega_1(T)\sigma_1^2(T) + \omega_2(T)\sigma_2^2(T)$ where ω_1, ω_2 are the fraction of pixels belonging to the two classes. Fig. 2c shows the result of adaptive thresholding of the filtered image in Fig. 2b. Due to the narrow artery widths and the use of local Otsu thresholding, vessel fragmentation is minimized.

3.3 Junction Extraction from Foreground Regions

To locate the junctions, we adopt Zhang and Suen[14] to skeletonize the binary thresholded image, as it is fast, simple, and outperformed other approaches we tried. The skeletonization of Fig. 2c is shown in Fig. 2d, and it preserves the main artery centerlines. By grouping connected components on interior pixels (non-junctions) of the skeletal image, we form skeletal curves $S_i = \{(x, y)\}$. The set of junctions is $J_m = (S_1, S_2, ...S_k)$ where the mth junction is the intersection of the incident skeletal curves $S_1, ..S_k$ to give a junction of degree k. Spurious

initial junctions which are pixels with at least two incoming curves, are clustered using this grouping and the centroidal pixel becomes a robust indicator of the actual junctions, as illustrated in the change from Fig. 2d to Fig. 2e.

3.4   Extracting Coronary Artery Segments

To extract the tubular regions from the skeletal curves, we look for boundary pixels on either side of the skeletal curve proceeding along surface normals at each skeletal point. The tubular boundary may have points out of order at turning points, particularly where there is ambiguity in surface normals (Fig. 2f). These out-of-order points are corrected by treating the chain of endpoints on either boundary as pairs of curves to be aligned using dynamic time warping [13]. The resulting artery fragments are shown in Fig. 2g for the skeletal curves in Fig. 2d.

Each coronary artery segment C_i is represented by an ordered set of skeletal points $\{(x, y, \delta1, \delta2, \theta)\}$ where $(x, y) \in S_i$ is a skeletal pixel on the skeletal curve S_i passing through the tubular segment, and $\delta1$, and $\delta2$ are the units along the surface normal θ at which the tubular boundary points are detected.

4   Feature Extraction from Coronary Angiograms

Given the skeletal representation, we next extract clinically meaningful features. The proposed features are supported by several clinical studies including those in the SYNTAX score and JACC011 guidelines[5,12].

4.1   Number of Significant Junctions

This gives an indication of the bushiness of arteries as diffuse arteries tend to have a larger number of junctions. This feature f_1 is simply recorded by the number of junctions J_m computed in the section above.

4.2   Thickness of Arteries

A blockage in the middle of the artery appears as a sudden change in the width of the artery. The average thickness of a coronary artery segment is given by $W_i = \frac{\sum_j (\delta2_j - \delta1_j)}{P}$ where the $\sum_j$ is over the P skeletal points. The range of thickness variation within tubular regions is given as $R_i = (\delta21_i^{max} - \delta21_i^{min})$, and gives an indication of stenosis. The distribution of thickness of arteries is given by the feature $f_2 = H(W_i)$ and $f_3 = H(R_i)$, where $H(W_i)$ and $H(R_i)$ are the histograms of the average thickness distribution and range respectively. The peaks in the histograms indicate the widths of dominant arteries and are useful in identifying the major coronary segments.

4.3   Number of Trifurcations

Trifurcations are useful to detect in cases when left main trifurcating coronary artery disease is present. Certain viewpoints (e.g. caudal) can cause trifurcation junctions to be detected, which is a good viewpoint descriptor. This feature is simply computed as $f_4 = \{J_m \, , \, \text{degree}(J_m) > 3\}$.

4.4   Tortuosity

Tortuosity is the number of curvature changes in the skeletal curves, measured using a histogram. The significant peaks in the histogram indicate the variation in tortuosity across the coronary artery segments. To estimate the curvature changes, we form a line segment approximation to the skeletal curve by recursively partitioning it at points of maximum deviation. These points of deviation are places where there is significant change in curvature. The tortuosity is normalized by taking the ratio of the curvature change points N_k over the total number of points N_i to give $T_i = \frac{N_k}{N_i}$. By repeating this over all curves, we form the tortuosity feature vector as a histogram over T_i as $f_5 = H(T_i)$.

4.5   Lengths of Artery Segments

The length of artery segments is important to assess early bifurcation of the left main coronary artery. Since the skeletal curves are available, this feature is easily computed by the pixel length of the skeletal curves and forming a histogram of it $f_6 = H(|S_i|)$, where $|S_i|$ is the length of skeletal curve S_i.

4.6   Lumen Variations

To measure the lumen variations, we sample the intensity in the original image along surface normals to the skeletal curve and average it at each point along the skeletal curve. The range in intensity variations is similarly normalized by the maximum intensity in the region and its histogram is feature $f_7 = H(I_i)$ where I_i is the normalized range of intensity variation in skeletal curve S_i.

5   Finding Similar Coronary Angiograms

By arranging the features into one long feature vector, each angiogram is represented by a vector F_c. Simple Euclidean distance comparisons between feature vectors is not sufficient to retrieve similar angiograms, both because of errors in vessel detection and the inherent variation in raw feature vectors.

Going past the Euclidean metric, we attempt to learn a distance metric so that vessels that are "similar" end up close to each other in feature space. Specifically, we adopted a recent work on a supervised metric learning method called Relevant Component Analysis (RCA)[11] as it has been shown to significantly improve clustering performance. RCA works by eliminating those dimensions that are

unlikely to be useful for classification using small subsets or chunklets of sample points. The chunklet covariance matrix is estimated as $C = \frac{1}{N}\sum_i \sum j(x_{ji} - m_j)(x_{ji} - m_j)^T$ where m_j denotes the mean of the j-th chunklet and x_{ji} the i-th element of the j-th chunklet. A whitening transformation is then associated with the covariance matrix $C\ W = C^{-\frac{1}{2}}$ to apply to the data points after dimensionality reduction[11].

To obtain the learning matrix, we first normalize all the feature vectors to be unit range and then annotate the features of a training set from distinct viewpoints, so that those from the same viewpoint and similar coronary anatomies are annotated with the same label. Using the resulting metric learned W, the distance between any two coronary angiograms is simply given by the Mahalanobis distance $(X2 - X1)^T W(X2 - X1)$. This distance is finally used to rank coronary angiogram images in a database using their respective feature vectors.

6   Results

From a collection of 1600 runs of X-ray angiography videos from 70 patients, we applied a keyframe detection method [13] to retain the top 10 key frames from each run, generating a ground truth test set of 600 images drawn across multiple patients, viewpoints and coronary arteries. The training set for metric learning was derived from another subset of keyframes chosen from the runs to depict distinct viewpoints such as anterior oblique projection, left anterior oblique, caudal and right coronary artery view. The training and testing was done on different sets of patients.

6.1   Accuracy of Junction Detection

First, trained experts manually counted the number of junctions observed in a set of about 130 left and right coronary images in 3 viewpoints. The spatial overlap with the automatically detected junctions is shown in Table 1. The 10% non-overlap in the spurious junctions is mostly from non-artery regions or from vessel overlaps and intersections (e.g. cross-overs) while all manually identified junctions were consistently detected. There is a large agreement between the manual and automatically found junctions. The lower overlap of RCA is due to the lower visibility of RCA over LCA in the X-ray images provided.

Table 1. Accuracy of junction detection in coronary arteries

Artery class	Number of Images	Avg. Manually detected Junctions per image	Automatically detected junctions	% Spatial Overlap
LCA	31	62.6	75.3	89.8%
LCA Caudal View	50	45.8	67.6	93.5%
RCA	50	24.3	37.5	74.3%

6.2   Accuracy of Coronary Artery Segment Extraction

First, trained experts marked major arteries using LabelMe[9] to trace the contours of the segments. We automatically detected tubular structures to extract the coronary segments and measured the spatial overlap between the two regions. Table 2 summarizes the average accuracy of the detection process in comparison with the ground truth for the three artery classes in our dataset. Here we define average overlap as the fraction of pixels that are in both manual and automatically extracted artery segments over the pixel area of the manual segments. Similarly, average non-overlap is defined the fraction of pixels in automatic regions that are not within manually indicated segments over the total number of white pixels in all manual and automatically detected regions. Due to the different normalization used, the two numbers need not add up to 100%. From this, we conclude that a large fraction of the artery regions are accurately detected.

Table 2. Accuracy of coronary artery segment detection

Artery class	Number of Images	Average Overlap	Average Nonoverlap
LCA	31	95.3%	12.4%
LCA Caudal View	50	93.8%	13.5%
RCA	50	86.2%	6.2%

6.3   Similarity Retrieval Performance

Using the learned distance matrix, we used query images from the 600 image test set to retrieve the top 10 most similar images from the same set. Fig. 3 shows a sample result with the query image in the top-left, and the top 5 similar images in ranked order ordered left to right, top to bottom. The retrieved images have similar topology, lumen variation, variation in artery thickness, and the same disease (left main). To evaluate precision and recall, we selected 200 images from the viewpoint set and asked trained experts to mark clinically similar images in the 600 image data set. We then measured the recall as the fraction of these images returned by the similarity ranking in the top K list while precision was
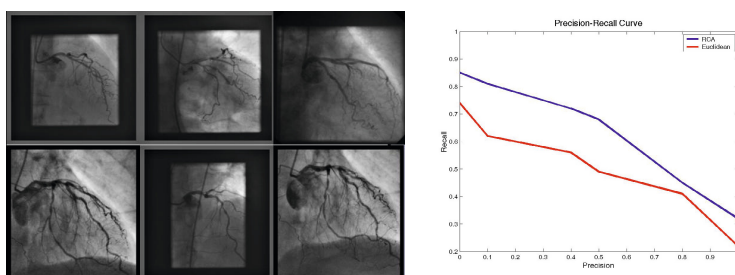
Fig. 3. Similarity retrieval of coronary angiograms. (a) Query angiogram image. (b)-(f) ranked order of matching coronary angiograms. (b) Precision-recall curve.

measured by the fraction of the top K matches returned that were relevant. The parameter K was varied to obtain the precision-recall curve. Fig. 3 shows the precision-recall curve using RCA-based metric learning in comparison to the Euclidean metric, indicating a large improvement in performance. In general, we found that the similarity retrieval preserved the identity of the arteries in the top 10 hits when the viewpoints were mixed in the dataset.

7   Conclusions

In this paper, we address for the first time, the problem of finding similar coronary angiograms using clinically meaningful features whose variation across patient population is learned using metric learning.

References

1. Chalopin, C., Finet, G., Magnin, I.E.: Modeling the 3D coronary tree for labeling purposes. Medical Image Analysis 5(4), 301–315 (2001)
2. Chen, T., Funka-Lea, G., Comaniciu, D.: Robust and Fast Contrast Inflow Detection for 2D X-ray Fluoroscopy. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part I. LNCS, vol. 6891, pp. 243–250. Springer, Heidelberg (2011)
3. Dehkordi, M.T., Sadri, S., Doosthoseini, A.: A review of coronary vessel segmentation algorithms. Journal of Medical Signals and Sensors 1(1) (2011)
4. 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)
5. Hsu, J.T.: Impact of calcification length ratio on the intervention for chronic total occlusions. Intl. Jl. of Cardiology, 135–141 (2011)
6. Kumar, R., Reina, A.V., Pfister, H.: Radon-like features and their application to connectomics. In: IEEE Workshop MMBIA (2010)
7. Lara, S., et al.: A semi-automatic method for segmentation of the coronary artery tree from angiography. In: ACM Symposium on Graphics, vol. 27(3), pp. 236–239 (2009)
8. Otsu, N.: A threshold selection method from gray-scale histogram. IEEE Trans. Syst., Man, Cybern. 9(1), 62–66 (1979)
9. Russell, B.C., et al.: LabelMe: a database and web-based tool for image annotation. International Journal of Computer Vision, 157–173 (2008)
10. Sato, Y., et al.: Three-dimensional multi-scale line filter for segmentation and visualization of curvilinear structures in medical images. In: Medical Image Analysis, pp. 143–168 (1998)
11. Shental, N., Hertz, T., Weinshall, D., Pavel, M.: Adjustment Learning and Relevant Component Analysis. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002, Part IV. LNCS, vol. 2353, pp. 776–790. Springer, Heidelberg (2002)
12. Sianos, G., et al.: The syntax score: and angiographic tool grading the complexity of coronary artery disease. EuroIntervention, 219–227 (2005)
13. Syeda-Mahmood, T., et al.: Automatic selection of keyframes from angiogram videos. In: ICPR (2010)
14. Zhang, T.Y., Suen, C.: A fast parallel algorithm for thinning digital patterns. Communications of ACM 27(3), 236–239 (1984)
15. Zhao, F., Bhotika, R.: Coronary artery tree tracking with robust junction detection in 3D CT angiography. In: ISBI, pp. 2066–2071 (2011)

Detection of Vertebral Body Fractures Based on Cortical Shell Unwrapping

Jianhua Yao[1], Joseph E. Burns[2], Hector Munoz[1], and Ronald M. Summers[1]

[1] Radiology and Imaging Sciences Department, Clinical Center, The National Institutes of Health, Bethesda, MD 20892
[2] Department of Radiological Sciences, University of California, Irvine School of Medicine

Abstract. Assessment of trauma patients with multiple injuries can be one of the most clinically challenging situations dealt with by the radiologist. We propose a fully automated method to detect acute vertebral body fractures on trauma CT studies. The spine is first segmented and partitioned into vertebrae. Then the cortical shell of the vertebral body is extracted using deformable dual-surface models. The extracted cortical shell is unwrapped onto a 2D map effectively converting a complex 3D fracture detection problem into a pattern recognition problem of fracture lines on a 2D plane. Twenty-eight features are computed for each fracture line and sent to a committee of support vector machines for classification. The system was tested on 18 trauma CT datasets and achieved 95.3% sensitivity and 1.7 false positives per case by leave-one-out cross validation.

1 Introduction

Assessment of trauma patients with multiple injuries, particularly in the setting of multiple trauma patients presenting to the hospital concurrently, can be one of the most clinically challenging situations dealt with by the radiologist. Traumatic injury of the spine is a subset of the spectrum of blunt trauma pathology, and is common and potentially devastating. Previous reports estimate the number of vertebral fractures each year in the United States at more than 140,000, with 19%-50% of fractures of the thoracolumbar spine associated with neurological injury [1]. Rapid and accurate assessment is essential for determination of an acceptable management algorithm, and delay in detection and management of spinal injuries can result in prolonged pain and suffering, or biomechanical disability.

Limited forays have been performed in prior work investigating computer-aided assessment of fractures, mainly in the realm of the detection of fractures on plain film radiographs for limited clinical circumstances [2]. Design of algorithms has been performed for simple assessment of anterior height loss of thoracolumbar vertebrae, and has recently reached the stage of clinical application limited to plain film radiograph lateral views of the spine [3]. There are also prior works assessing for fractures based on detected global geometric deformities of the vertebral bodies (compression deformities), rather than direct detection of fracture lines in the vertebrae [4] [5]. Analysis of the complex structure of the spine on cross sectional CT images for direct visualization of fractures is a novel topic of clinical importance, and is the goal and focus of this investigation. To the best of our knowledge, our

investigation is the first to target computer-aided detection of fracture-associated bone discontinuities in vertebral bodies using CT images.
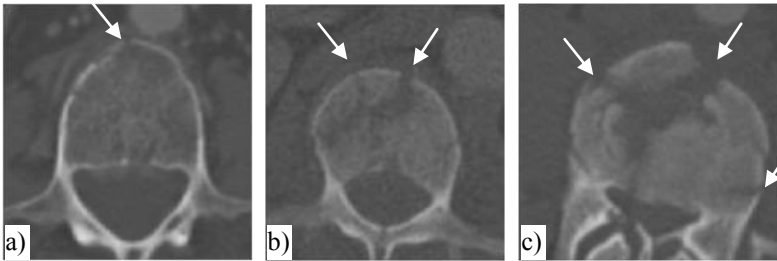
Fig. 1. Examples of vertebral body fractures (arrows) on axial CT images in three different patients

Figure 1 shows examples of vertebral body fractures in axial CT images. Fractures occur in myriad patterns, and vertebrae exhibit complex geometry, posing a complex mathematical problem. Our idea is to directly detect the osseous fracture line involvement of the anterior, posterior, and lateral cortices of the vertebral body. Thus, we propose a novel method to convert the complex 3D detection problem into a 2D pattern recognition problem by unwrapping the cortical shell of the vertebral body.

2   Methods

Our method is summarized as follows. Given a spine CT data set, the spinal column is first extracted and partitioned into individual vertebrae. The cortical shell of vertebral body is then segmented using deformable dual-surface models. After that, the cortical shell is unwrapped onto a 2D plane. Pattern recognition techniques are then applied to detect fracture lines on the unwrapped cortical shell. These detections are then re-projected back to 3D space and quantitative features are computed. Finally, the detections are passed to a committee of support vector machines for classification.

2.1   Spinal Column Segmentation and Partitioning

First, thresholding and connected component analysis are conducted to obtain the initial spine segmentation. The spinal canal is then extracted using a watershed algorithm and a directed acyclic graph search. Next, curved planar reformation is computed along the centerline of the spinal canal to partition the spinal column into individual vertebrae. Details of the automated spinal column extraction and partitioning can be found in [6].

2.2   Cortical Shell Segmentation

In a fractured vertebral body, the cortical shell is often damaged with cracks or broken into disconnected components, posing challenges for its segmentation. Concentric ring approach can't segment cortical shell correctly. We propose a deformable dual-surface model to extract both the exterior and interior (periosteal and endosteal) surfaces of the cortical shell.
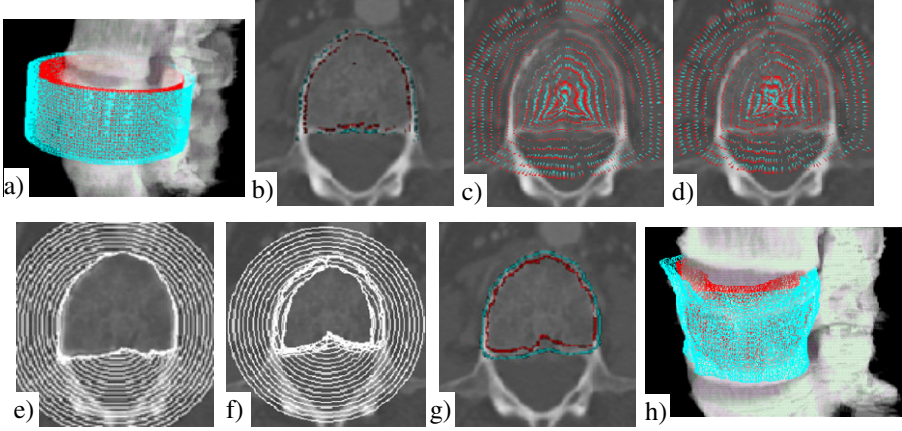
Fig. 2. Cortical shell segmentation. The original image is in Figure 1a. a) Initial models; b) Potential boundary map (R_E, R_I), cyan: R_E, red: R_I; c) Potential force for exterior surface $P(S_E)$; d) Potential force for interior surface $P(S_I)$. Maps are down-sampled for clarity. Force direction points from red to cyan; e) Evolution of exterior surface; f) Evolution of interior surface; g) Results of dual-surface segmentation; and h) 3D visualization.

After the spine is extracted, a local cylindrical coordinate system is established for each vertebral body. An initial dual surface is placed in the center of the vertebral body. The height is set as the distance between the superior and inferior end plates and the radius is estimated as twice of the average width of the vertebral body (Figure 2a). The surface is constructed as a triangular mesh, where the vertices are evenly spaced. The surface can be represented as $r = S(z, \varphi)$ in the cylindrical coordinate system, where z is the height along the axis, φ is the azimuth angle, and r is the radial distance. r is uniform at every point on the initial exterior (S_E) and interior (S_I) surfaces. The resolution of the surface mesh is set to be the same as the CT image.

The deformable dual-surfaces [7] are driven by internal forces, image potential forces, and constraints between the exterior and interior surfaces. The energy functional for the dual-surface is written as,

$$
\begin{aligned}
\mathrm{E}(S_E) &= w_i I(S_E) + w_p P(S_E) + w_c C(S_E, S_I) \\
\mathrm{E}(S_I) &= w_i I(S_I) + w_p P(S_I) + w_c C(S_E, S_I)
\end{aligned}
\tag{1}
$$

where S_E and S_I are exterior and interior surfaces, $I(S)$ is the internal force, $P(S)$ is the image potential force, $C(S_E, S_I)$ is the constraint between the two surfaces, and w_i, w_p and w_c are weights for the three forces. The internal forces keep the surface smooth and continuous, which can be written as,

$$
I(S(z,\varphi)) = \int \left(\alpha \left| \frac{\partial S(z,\varphi)}{\partial z} + \frac{\partial S(z,\varphi)}{\partial \varphi} \right| + \beta \left| \frac{\partial^2 S(z,\varphi)}{\partial z^2} + \frac{\partial^2 S(z,\varphi)}{\partial \varphi^2} + \frac{\partial^2 S(z,\varphi)}{\partial z \partial \varphi} \right| \right) dz d\varphi
\tag{2}
$$

The first order derivative discourages the stretching and the second order derivative discourages the bending of the surfaces. α and β are weights and set to 1.

A directional gradient in the cylindrical coordinate system is applied to compute the potential image. For a point (z, φ, r) on the image, the directional gradient is

$$\nabla\bar{G}(z,\varphi,r) = \frac{\partial G(z,\varphi,r)}{\partial r} \tag{3}$$

where G is the grayscale image. For every direction defined by (z, φ) (at level z, angle φ), we search for the maximum of a pair of positive and negative directional gradients to be used as the potential boundary for exterior and interior surfaces, i.e.,

$$\begin{pmatrix} R_E(z,\varphi) \\ R_I(z,\varphi) \end{pmatrix} = \underset{\substack{r_E, r_I \\ st, r_I < r_E, \varepsilon_1 < r_E - r_I < \varepsilon_2}}{\arg\max} \left(\left| \nabla\bar{G}^+(z,\varphi,r_E) \right| + \left| \nabla\bar{G}^-(z,\varphi,r_I) \right| \right) \tag{4}$$

Here $\nabla\bar{G}^+$ and $\nabla\bar{G}^-$ represent positive and negative directional gradient respectively, and ε_1 and ε_2 are the minimum and maximum cortical shell thickness. Due to the image noise and other anatomical structures near the vertebral body, (R_E, R_I) may become stuck at false edges. In order to eliminate outliers, we fit a Bezier function for R_E over the domain of (z, φ). Those (R_E, R_I) pairs that are far away from the Bezier function are excluded. Figure 2b shows an example of (R_E, R_I) map superimposed on an image slice. The distance to the (R_E, R_I) map is then used to derive the potential force for the dual surfaces, which can be formulated as,

$$P(S_E(z,\varphi)) = \left\| S_E(z,\varphi) - R_E(z,\varphi) \right\|$$
$$P(S_I(z,\varphi)) = \left\| S_I(z,\varphi) - R_I(z,\varphi) \right\| \tag{5}$$

where $\| \ \|$ is the Euclidean distance. Figures 2c and 2d show the potential forces $P(S_E)$ and $P(S_I)$.

The constraint between the dual surfaces is the thickness of the cortical shell. We assume that the thickness should be continuous over the cortical shell, and use the following function for the thickness regulation,

$$C(S_E, S_I) = \int \left(\left| \frac{\partial(S_E(z,\varphi) - S_I(z,\varphi))}{\partial z} \right| + \left| \frac{\partial(S_E(z,\varphi) - S_I(z,\varphi))}{\partial \varphi} \right| \right) dz d\varphi \tag{6}$$

The weights for different forces (w_i, w_p and w_c) in equation 1 are kept constant throughout the evolution. Since the potential force becomes smaller when closer to the boundaries (see Fig. 2c and 2d), the internal force and thickness constraint will play a bigger role upon convergence. Figures 2e and 2f show the evolution of the exterior and interior surfaces. Figures 2g and 2h show the final segmentation results on one 2D slice and in 3D space.

2.3   Cortical Shell Unwrapping

The unwrapping of the cortical shell is based on the cylindrical coordinate system. We map the 3D cortical shell onto the 2D space of (z, φ). The unwrapping process is,

$$U(z,\varphi) = \frac{1}{\left\| S_E(z,\varphi) - S_I(z,\varphi) \right\|} \int_{S_I(z,\varphi)}^{S_E(z,\varphi)} G(z,\varphi,r) dr \tag{7}$$

here G is the image intensity. Essentially, we project the mean intensity of the cortical shell onto a 2D map. The mapping is one-to-one: any point on the unwrapped map has a corresponding point on the 3D cortical shell. Figure 3a shows an example of the cortical shell unwrapped map. The horizontal axis is φ and the vertical axis is z. Axis φ starts from the center of the spinal canal (detected in section 2.1) and spans 360°, and axis z goes from the inferior to the superior endplates.
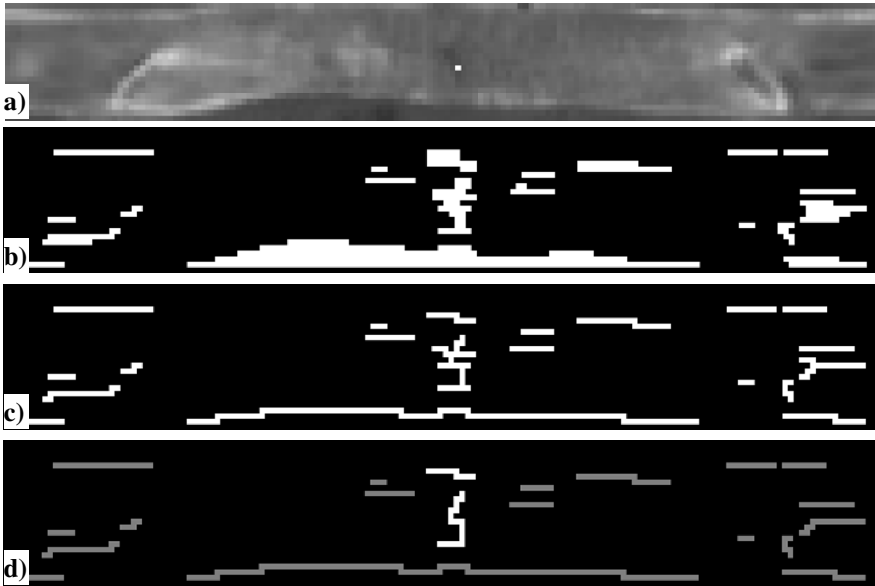
Fig. 3. Cortical shell unwrapping and fracture line detection. Original CT image is in Figure 1a. a) Unwrapped cortical shell map (bright dot indicates the projected fracture site marked by an expert on the original CT); and results after b) Adaptive filtering; c) Skeletonization; d) Pruning (the bright line is the true fracture line).

2.4   Fracture Line Detection

Fracture lines on the cortical shell appear as gaps or discontinuities on the unwrapped map (Figure 3a). Detecting discontinuities on the map is a relatively simple 2D pattern recognition problem compared to the complex 3D fracture detection problem, similar to road crack detection in computer vision applications [8]. We adopt a multi-scale adaptive filtering method to detect cracks on the unwrapped map. Two assumptions are applied: 1) a crack is darker than the background (normal cortical shell); and 2) a crack is composed of a set of connected segments with different orientations and limited width. We define the crack filter as a rectangle function,

$$f(x) = \begin{cases} -1 & x \in [-T/2, T/2] \\ 1 & Elsewhere \end{cases}$$

$$(8)$$

where T is the width of the crack and also the scale of the filter. We convolve the unwrapped map $U(z, \varphi)$ with $f(x)$ of different scales (by varying T) at multiple

orientations ([*0, π/4, π/2, 3π/4*]). The outputs of all filters are merged and used as the initial detection for the fractured regions (Figure 3b). The merging is additive and a connected component analysis is conducted to obtain detections.

We then apply a Hilditch thinning algorithm [9] to skeletonize the fracture region (Fig 3c). After that, the branches on the skeleton are pruned [10] so that only the longest path remains, and is detected as one potential fracture line (Fig 3d).

2.5   Feature Extraction and Classification

Many false positives remain after the filtering. We extract a set of 28 quantitative features to differentiate true fracture lines from false detections. The features for each fracture line can be roughly partitioned into four categories: location (e.g. circumferential angle, distance, and orientation), shape (e.g. width, thickness, aspect ratio), intensity (e.g. intensity, contrast) and attributions of its associated vertebral body (e.g. height, radius, and intensity). Due to the page limit, here we only list the formula for a few features (f_1:average width, f_2:average thickness, f_3:average intensity, and f_4:average interior intensity),

$$f_1 = \frac{1}{\|\Omega_s\|} \sum_{(z,\varphi) \in \Omega_s} d(z,\varphi) \tag{9}$$

$$f_2 = \frac{1}{\|\Omega_A\|} \sum_{(z,\varphi) \in \Omega_A} \left(S_E(z,\varphi) - S_I(z,\varphi) \right) \tag{10}$$

$$f_3 = \frac{1}{\|\Omega_A\|} \sum_{(z,\varphi) \in \Omega_A} U(z,\varphi) \tag{11}$$

$$f_4 = \frac{1}{\|\Omega_A\|} \sum_{(z,\varphi) \in \Omega_A} \int_0^{S_I(z,\varphi)} \frac{G(z,\varphi,r)}{S_I(z,\varphi)} dr \tag{12}$$

where Ω_A is the set of points in the detected fracture region, Ω_S is the skeleton of Ω_A, and d is the distance from a skeleton point to its closest boundary point. The features are computed from both the 3D CT data and the 2D unwrapped map.

A committee of support vector machines (SVM) [11] is trained to classify the detections into true fracture lines or false ones. The training was based on reference standard of fractures marked by an expert. A forward stepwise feature selection procedure was conducted to form a seven-member committee. Each committee member had three characteristic features (features may overlap among committee members). Ten-fold cross validation was employed to evaluate the performance.

3   Data Sets and Experimental Results

Our cohort includes 18 trauma patients admitted to UC Irvine Medical Center between June 2009 and July 2010. The mean patient age was 51±11 yrs (18-86yrs). There were

13 men and 5 women. All patients were scanned on a Siemens Sensation 64 scanner. The scanning parameters were: 2mm slice thickness, 120 kvp, no intravenous contrast administration, and convolution kernel B40f (16 patients) or B60f (2 patients). The CT data covered the thoracic and lumbar spines, and included 14 vertebrae on average. An expert radiologist examined the cases and manually marked the fracture sites. Ten patients were positive for vertebral body fractures. The total number of spatially distinct fracture sites was 21, among the 10 patients. The remaining 8 patients had no evidence of vertebral body fracture. The average running time is 5.6 minutes.
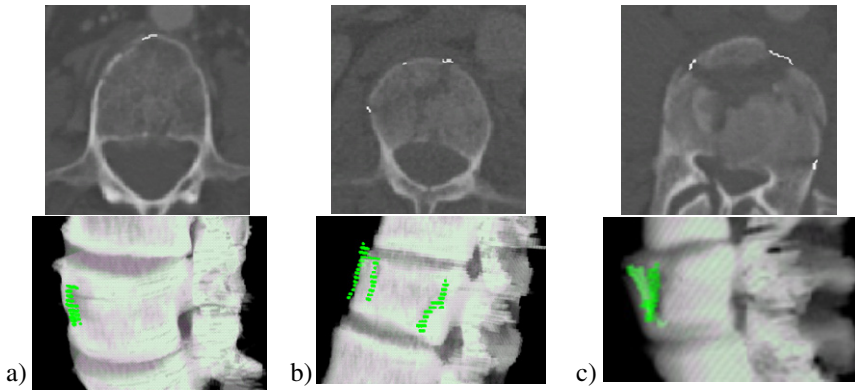
Fig. 4. Examples of detected fracture lines. Original images are in Figure 1. First row: 2D view; second row: 3D view.
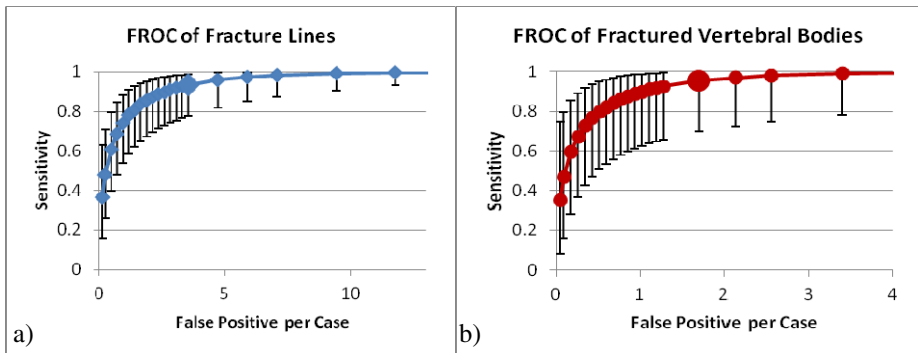
Fig. 5. FROC analysis. Big markers indicate operating points.

Figure 4 shows examples of detected fracture lines. Our method successfully detected both nondisplaced and simple fractures (Fig. 4a, 4b), as well as burst fractures (Fig. 4c). Figure 5a shows the FROC curve of fracture line detection. We also evaluated the performance on a per-vertebra basis, by merging all detections in one vertebral body into one "detection" (only keeping the detection with highest SVM value) and report the number of vertebrae having any fracture (FROC in Figure 5b). The system achieved 92.7% sensitivity (95% confidence interval: [76.7%, 98.5%]) at 3.3 false fracture sites per patient, and 95.3% sensitivity (95% CI: [72.6%, 99.9%]) at

1.7 false fractured vertebrae per patient. The FROC analysis was conducted using ROCKIT (*http://xray.bsd.uchicago.edu/krl/KRL_ROC/software_index6.htm*).

The etiology of common false positives includes costovertebral junctions, partial volume averaging of vertebral disks and nutrient vessel foramen.

4   Discussion

Our method converts a complex 3D fracture detection problem into a simpler 2D pattern recognition problem and achieves high sensitivity and specificity. The system may serve as a shadow reader to assist radiologists and has the potential to increase detection rates for spinal fractures at early points after occurrence, allowing appropriate management and preventing secondary complications.

Future work will include detection of fractures in spinous and transverse processes. We will also incorporate rib segmentation in the system to reduce the number of false positives at the costovertebral junctions.

References

1. Smith, M., et al.: The Reliability of Nonreconstructed Computerized Tomographic Scans of the Abdomen and Pelvis in Detecting Thoracolumbar Spine Injuries in Blunt Trauma. J. Bone Joint Surg. Am. 91, 2342–2349 (2009)
2. He, J.C., Leow, W.-K., Howe, T.S.: Hierarchical Classifiers for Detection of Fractures in X-Ray Images. In: Kropatsch, W.G., Kampel, M., Hanbury, A. (eds.) CAIP 2007. LNCS, vol. 4673, pp. 962–969. Springer, Heidelberg (2007)
3. Guglielmi, G., et al.: Assessment of osteoporotic vertebral fractures using specialized workflow software for 6-point morphometry. European Journal of Radiology 70(1), 142–148 (2009)
4. Ghosh, S., et al.: Automatic Lumbar Vertebra Segmentation from clinical CT for Wedge Compression Fracture Diagnosis. In: SPIE Medical Imaging (2011)
5. Yao, J., et al.: Quantitative Vertebral Compression Fracture Evaluation Using a Height Compass. In: SPIE Medical Imaging (2012)
6. Yao, J., O'Connor, S.D., Summers, R.M.: Automated Spinal Column Extraction and Partitioning. In: IEEE ISBI, Arlington, VA (2006)
7. Xu, C., Pham, D.L., Prince, J.L.: Medical Image Segmentation Using Deformable Models. In: Sonka, M., Fitzpatrick, J.M. (eds.) Handbook of Medical Imaging, Medical Image Processing and Analysis, vol. 2, pp. 129–174. SPIE (2000)
8. Chambon, S., et al.: Road crack extraction with adapted filtering and Markov model-based segmentation. In: International Joint Conference on Computer Vision Theory and Applications, Angers, France (2010)
9. Zhang, T.Y., Suen, C.Y.: A Fast Parallel Algorithms For Thinning Digital Patterns. Communication of the ACM 27(3), 236–239 (1984)
10. Bai, X., Latecki, L.J., Liu, W.-Y.: Skeleton Pruning by Contour Partitioning with Discrete Curve Evolution. IEEE Transactions on Pattern Analysis and Machine Intelligence 29(3), 1–14 (2007)
11. Cristianini, N., Taylor, J.S.: An Introduction to Support Vector Machines. Cambridge University Press (2000)

Multiscale Lung Texture Signature Learning Using the Riesz Transform

Adrien Depeursinge[1,2], Antonio Foncubierta–Rodriguez[1],
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
http://medgift.hevs.ch/

Abstract. Texture–based computerized analysis of high–resolution computed tomography images from patients with interstitial lung diseases is introduced to assist radiologists in image interpretation. The cornerstone of our approach is to learn lung texture signatures using a linear combination of N–th order Riesz templates at multiple scales. The weights of the linear combination are derived from one–versus–all support vector machines. Steerability and multiscale properties of Riesz wavelets allow for scale and rotation covariance of the texture descriptors with infinitesimal precision. Orientations are normalized among texture instances by locally aligning the Riesz templates, which is carried out analytically. The proposed approach is compared with state–of–the–art texture attributes and shows significant improvement in classification performance with an average area under receiver operating characteristic curves of 0.94 for five lung tissue classes. The derived lung texture signatures illustrate optimal class–wise discriminative properties.

Keywords: Texture analysis, Riesz, steerability, interstitial lung diseases, high–resolution computed tomography, computer–aided diagnosis.

1 Introduction

Objective assessment of texture information is a difficult task in radiology [1]. Texture 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 [2]. Early detection of diffuse disease conditions requires to analyze very subtle changes in texture properties of the image, where computerized image processing proved to significantly outperform clinical experts [1]. The various appearances of lung tissue affected by interstitial lung diseases (ILD) in high–resolution computed tomography (HRCT) are best characterized in terms of texture properties [3]. Differentiation of these patterns is regarded as difficult even for experienced radiologists. Consequently, several studies investigated the potential of computerized classification of the lung parenchyma to assist radiologists in HRCT interpretation [4–7]. To ensure the success of such a system, the

ability of the image attributes to encode the subtle texture signatures associated with the various lung tissue types is crucial. In particular, localized quantification of orientations and scales is known to be relevant for texture discrimination [8, 9]. Whereas most of the image analysis approaches to texture feature extraction are based on the characterization of these two affine properties, all require arbitrary sampling of at least one of these two parameters (e.g., grey–level co–occurrence matrices (GLCMs) [9], run–length matrices (RLE), local binary patterns (LBP) [8], and non–steerable Gabor or Gaussian filterbanks [10]). The choice of scales and orientations has a direct impact on system performance and is difficult since these properties vary for each image pixel.

In this article, we introduce a novel texture analysis approach allowing translation invariance as well as scale and rotation covariance with infinitesimal precision. It extends our previous work [11] by using support vector machines (SVM) to learn the respective relevance of multiscale Riesz components. Class–wise texture signatures are then obtained from linear combinations of the latter, allowing for visual assessment of the learned texture patterns.

2   Material and Methods

2.1   Dataset

A publicly available dataset of 85 ILD cases with annotated HRCT images is used to evaluate our approach [12]. Expert annotations were carried out in collaboration by two radiologists with 15 and 20 years of experience in CT imaging. The slice thickness is 1mm and the inter–slice distance is 10mm. The images were acquired with two imaging devices at the Radiology Service of the University Hospitals of Geneva: a Philips Mx8000 IDT 16 CT Scanner and a General Electric HiSpeed CT. The five lung tissue classes encountered in most ILDs were chosen as lung texture classes: healthy (H), emphysema (E), ground glass (G), fibrosis (F) and micronodules (M). In each annotated slice, 2D hand–drawn regions of interests (ROIs) are divided into 32×32 square blocks. The visual appearance of the lung texture classes and their distribution are detailed in Fig. 4.

2.2   Multiscale Lung Texture Signature Learning

The cornerstone of our approach to multiscale lung texture signature learning is to use the structural risk minimization principle to derive class–wise texture prototypes from the Riesz transform. The obtained class–wise texture signature has optimal discriminative properties for a given one–versus–all (OVA) classification task. The Riesz transform yields steerable filterbanks and commutes with translation, scaling or rotation [13]. The richness of the filterbank is controlled by the order N of the Riesz $\mathcal{R}$ transform as:

$$\widehat{\mathcal{R}^{(n_1,n_2)}f}(\boldsymbol{\omega}) = \sqrt{\frac{n_1 + n_2}{n_1!n_2!}} \frac{(-j\omega_1)^{n_1}(-j\omega_2)^{n_2}}{||\boldsymbol{\omega}||^{n_1+n_2}} \hat{f}(\boldsymbol{\omega}),\tag{1}$$
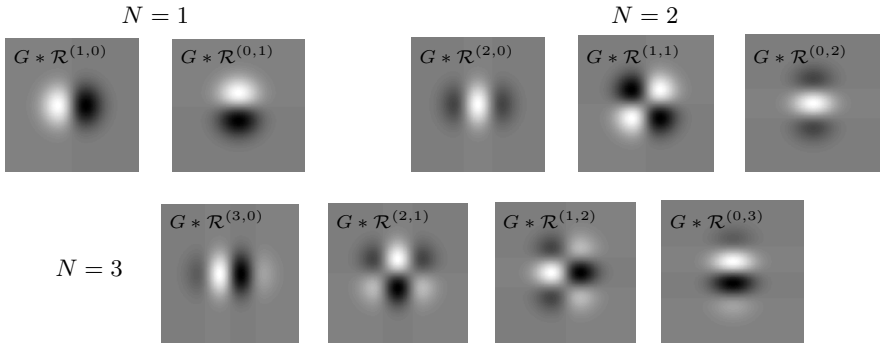
Fig. 1. Riesz filterbanks for N=1,2,3

for all combinations of (n_1, n_2) with $n_1 + n_2 = N$ and $n_{1,2} \in \mathbb{N}$. The vector $\boldsymbol{\omega}$ is composed of $\omega_{1,2}$ corresponding to the frequencies in the two image axes, and $\hat{f}(\boldsymbol{\omega})$ denotes the Fourier transform of $f(\boldsymbol{x})$. The Riesz transform yields $N{+}1$ distinct components behaving as N–th order directional differential operators. Riesz components $\mathcal{R}^{(n_1,n_2)}$ convolved with isotropic Gaussian kernels $G(\boldsymbol{x})$ for N=1,2,3 are depicted in the spatial domain in Fig. 1. Multiscale versions of the filterbanks are obtained by coupling the Riesz transform with Simoncelli's multi–resolution framework based on isotropic band–limited wavelets [14]. Four scales $s_i = \{1, \ldots, 4\}$ with a dyadic progression are used to cover the Nyquist domain. The Riesz wavelet filterbanks are steerable, which means that the response of each component $G * \mathcal{R}^{(n_1,n_2)}$ rotated by an arbitrary angle θ can be derived analytically from a linear combination of the responses of all components of the filterbank [15, 13]. This property enables rotation covariance of the proposed texture descriptors with infinitesimal angular precision. To ensure that the distribution of the directional information is normalized among the Riesz components for any rotation of the texture patterns, each components are all locally aligned to maximize the response of $G*\mathcal{R}^{(N,0)}$ at the finest scale, which is carried out analytically and proved to improve lung texture classification performance in [11]. This enables rotation invariance of the texture descriptors without discarding precious orientation information, which is often lost when using isotropic detectors [7, 6] or when averaging the responses of multi–oriented features as it is commonly carried out for GLCMs, RLEs and Gaussian filterbanks [5]. To summarize, the Riesz wavelets benefit from the steerability property while enabling much richer feature extraction than rotated filterbanks and classical steerable filterbanks [16]. Therefore, it allows multiscale and multi–directional image analysis with infinitesimal spatial and angular precision[1].

In order to optimally exploit the richness of the feature detectors encompassed in the multiscale Riesz components for a given texture classification task, an

---

[1] In the discrete domain, the spatial and angular precisions are determined by Nyquist.

appropriate weighting scheme of the energy of the responses of the multiscale Riesz components is required. The goal is to build an optimal texture signature Γ_c^N of the class c (versus all) from a linear combination of the multiscale Riesz components as:

$$\Gamma_c^N = w_1 \left(G * \mathcal{R}^{(N,0)} \right)_{s_1} + w_2 \left(G * \mathcal{R}^{(N-1,1)} \right)_{s_1} + \cdots + w_{4N+4} \left(G * \mathcal{R}^{(0,N+1)} \right)_{s_4}. \quad (2)$$

l_1–norm support vector machines (SVM) are used to find the optimal weights $\boldsymbol{w}^{\mathrm{T}} = (w_1 \ldots w_{4N+4})$ in the sense of structural risk minimization [17] as:

$$\min_{\boldsymbol{w}, \xi, b} \left\{ \frac{||\boldsymbol{w}||_1^2}{2} + C \sum_{i=1}^n \xi_i \right\} \quad \text{subject to} \quad y_i(K(\boldsymbol{w}, \boldsymbol{x_i}) - b) \geq 1 - \xi_i, \quad \xi_i \geq 0. \quad (3)$$

where ξ is the slack variable of the soft margin, C is the cost of the errors, $\boldsymbol{x_i}$ are the texture instances $i = 1 \ldots n$ expressed in terms of the energy of the Riesz components, and y_i are the corresponding labels. $K(\boldsymbol{x_i}, \boldsymbol{x_j})$ is a Gaussian kernel as: $K(\boldsymbol{x_i}, \boldsymbol{x_j}) = \exp(\frac{-||x_i - x_j||_1^2}{2\sigma^2})$. The contribution of each Riesz component is determined by the weight that its corresponding energy level received in Eq. (3). For multiclass classification with N_c classes, the OVA approach is used. The model with the highest decision level for the positive class determines the final class c_{max} as: $\max_{c \in \{1, \ldots, N_c\}} \{K(\boldsymbol{w_c}, \boldsymbol{x_i}) - b\}$. The global workflow of the proposed approach for lung texture signature extraction and classification is summarized in Fig. 2.
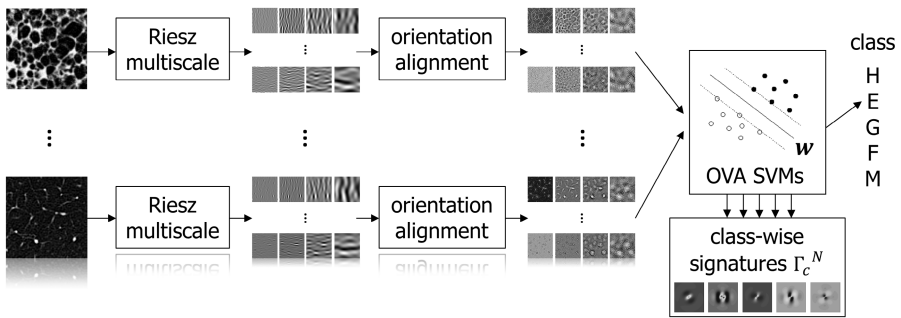
Fig. 2. Flow chart of lung texture signature learning and classification

3   Results

The proposed methods are evaluated both qualitatively and quantitatively on artificial and real lung textures. The principle of multiscale texture signature learning is first demonstrated on artificial data, where scale and rotation covariance are investigated in Fig. 3. All artificial texture patterns are containing noise and their signatures are learned when confronted to white noise. The first two
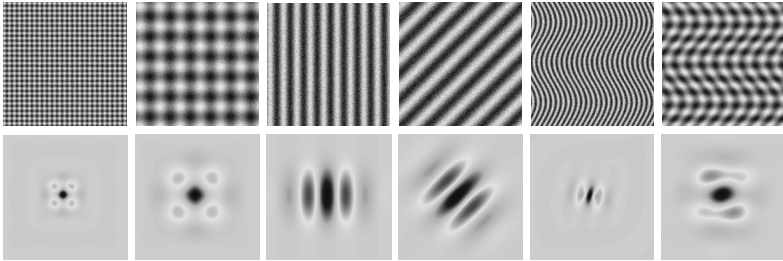
Fig. 3. Lower row: multiscale texture signatures Γ_c^8 of the upper row for $N{=}8$

healthy	emphysema	ground glass	fibrosis	micronodules
$\Gamma_{\text{healthy}}^4$	$\Gamma_{\text{emphysema}}^4$	$\Gamma_{\text{ground glass}}^4$	$\Gamma_{\text{fibrosis}}^4$	$\Gamma_{\text{micronodules}}^4$
3011 blocks, 7 patients.	407 blocks, 6 patients.	2226 blocks, 32 patients.	2962 blocks, 37 patients.	5988 blocks, 16 patients.
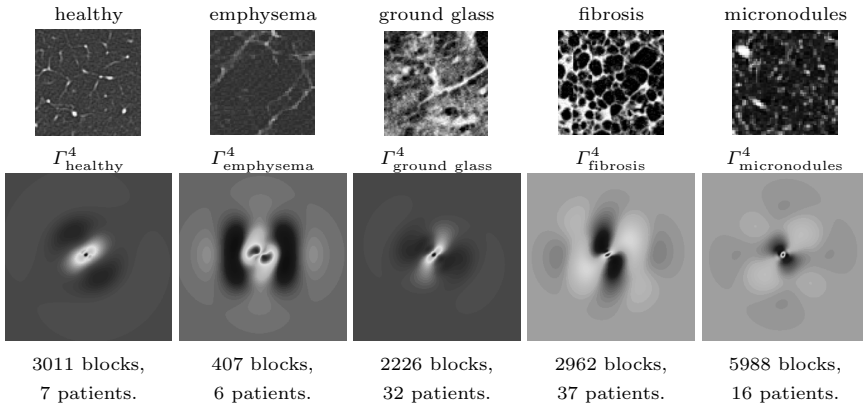
Fig. 4. Distribution of the texture classes and visual appearance of the class–wise multiscale lung texture signatures Γ_c^4

signatures are learned from two checkerboards with various scales. The distribution of the weights $\boldsymbol{w}$ for the scales $\{s_1, \ldots, s_4\}$ are $\{0.1\%, 18.5\%, 81.1\%, 0.3\%\}$ for the small scale and $\{2.3\%, 3.9\%, 14\%, 79.8\%\}$ for the large scale checkerboard. The rotation covariance is demonstrated with oriented stripes in the third and fourth columns of Fig. 3. Robustness to non–rigid transformations is illustrated with deformed versions of the stripes and checkerboard in the last two columns.

The visual appearance of the five lung tissue classes and the corresponding learned class–wise texture signatures over the entire dataset in OVA configurations are shown in Fig. 4. Fig. 5 shows the receiver operating characteristic (ROC) analysis of the classification performance of the proposed methods over the 85 folds of a leave–one–patient–out cross–validation. We compared our approach with two commonly used lung texture feature sets: LBPs [4] and GLCMs combined with RLEs [5]. We optimized the parameters of each approach using an exhaustive grid search. A radius $R \in \{1, 2\}$ pixels and a number of samples $P \in \{8, 16\}$ are used for LBPs, according to [4]. For GLCMs and RLEs, distances of $\{1, 2, \ldots, 5\}$ are used and the texture measures from [9] are averaged across orientations of $\{0°, 45°, 90°, 135°\}$. A grey–level reduction of 8 levels obtained
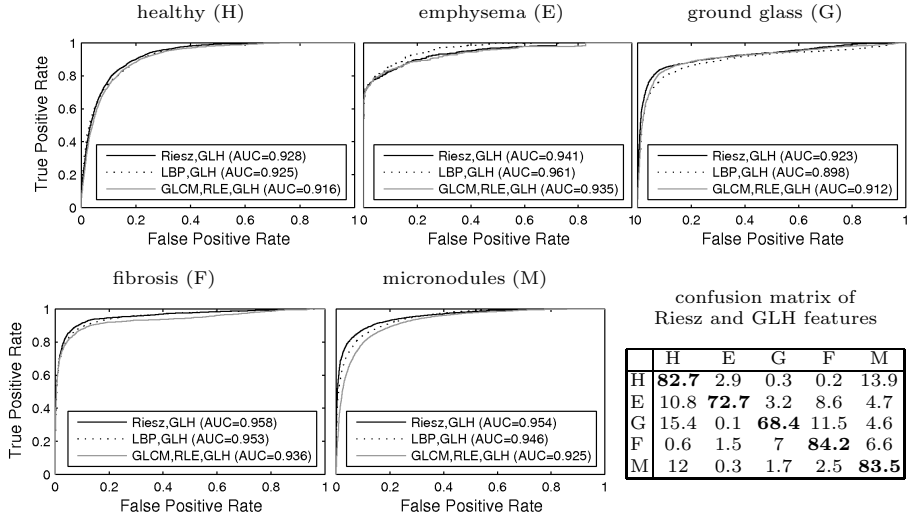
Fig. 5. ROC analysis for the various texture analysis approaches and confusion matrix. $N = 4$ for all Riesz features.

best performance when compared to 16 and 32. All approaches are combined with 22 grey level histogram (GLH) bins in [-1050;600] Hounsfield Units (HU) and the percentage of air pixels with values $\leq$-1000 HU. Best area under ROC curves (AUC) are of 0.941, 0.936 and 0.925 for Riesz (N=4), LBPs (R=1, P=8) and GLCMs with RLEs, respectively.

4   Discussions and Conclusions

We propose a novel texture analysis method to learn multiscale texture signatures based on Riesz wavelets and SVMs, which is rotation and scale covariant. A pixel–wise alignment of the Riesz templates ensures the normalization of the distribution of the directional information over the Riesz components, which allows both inter–instance rotation invariance and intra–instance rotation covariance, similarly to rotation–invariant LBPs [8]. The important scales and orientations are learned based on the structural risk minimization principle and therefore do not need a priori assumptions, which is an advantage when compared to other state–of–the–art texture features such as GLCMs, RLEs and LBPs. Linear combinations of multiscale Riesz components allow discovering class–wise important discriminatory patterns and visual analysis of their relevance. The multiscale texture signatures shown in Fig. 3 demonstrate the ability of our approach to characterize texture patterns with multiple and varying scales and orientations. A relative robustness to non–rigid transformations is also observed. The lung texture signatures depicted in Fig. 4 are showing important class–wise discriminative properties. $\Gamma^4_{\mathrm{fibrosis}}$ clearly resembles the fibrosis patterns characterized by air bubbles surrounded by high–density walls of collagen. The same is true for

$\Gamma^4_{\text{micronodules}}$, where the micronodule detector is clearly visible with a high peak in the center of the signature for relatively small scales. $\Gamma^4_{\text{healthy}}$ and $\Gamma^4_{\text{ground glass}}$ are found to be similar, which is coherent to the definition of ground glass characterized by a diffuse increased opacity, where the bronchovascular structures remain visible. $\Gamma^4_{\text{healthy}}$ and $\Gamma^4_{\text{ground glass}}$ are therefore implementing a hybrid ridge and peak detector corresponding to the projections of the bronchovascular structures in 2D HRCT slices. Emphysema patterns are the result of the destruction of lung tissue, which is replaced by air. This process does, therefore, not engender the typical texture signature that our method aims to learn. LBPs seem to better encode the transitions between air and parts of remaining tissue, which shows the potential of combining Riesz and LBP. ROC analysis of the classification performance of the texture analysis approaches reveals an excellent average performance AUC=0.94 for the proposed approach, based on realistic data and methodology. It outperforms LBPs and GLCMs combined with RLEs in terms of overall classification performance with high statistical significance based on a 1–tailed paired T–test: $p = 4.75 \times 10^{-20}$ for Riesz versus LBPs and $p = 5.59 \times 10^{-43}$ for Riesz versus GLCMs combined with RLEs. This performance suggests that it can provide valuable assistance in the difficult task of texture analysis of lung tissue patterns in clinical routine with high reliability. In future work, class–wise feature combination and selection among various Riesz orders and other texture features such as LBPs, GLCMs and RLEs will be investigated using SVM–based recursive feature elimination. We are also currently extending our approach to three dimensions. It is expected to provide even better results, since the number of possible scales and orientations increases exponentially in 3D. A priori knowledge on their organization in 3D is difficult to obtain, because textures existing in more than two dimensions cannot be fully visualized by humans.

References

1. Wagner, R.F., Insana, M.F., Brown, D.G., Garra, B.S., Jennings, R.J.: Texture discrimination: radiologist, machine and man. In: Blakemore, C., Adler, K., Pointon, M. (eds.) Vision, pp. 310–318. Cambridge University Press (1991)
2. Tourassi, G.D.: Journey toward computer–aided diagnosis: Role of image texture analysis. Radiology 213(2), 317–320 (1999)
3. Webb, W.R., Müller, N.L., Naidich, D.P.: High–Resolution CT of the Lung. Lippincott Williams & Wilkins, Philadelphia (2001)
4. Sørensen, L., Shaker, S.B., De Bruijne, M.: Quantitative analysis of pulmonary emphysema using local binary patterns. IEEE Transactions on Medical Imaging 29(2), 559–569 (2010)
5. Park, Y.S., Seo, J.B., Kim, N., Chae, E.J., Oh, Y.M., Lee, S.D., Lee, Y., Kang, S.H.: Texture–based quantification of pulmonary emphysema on high–resolution computed tomography: Comparison with density–based quantification and correlation with pulmonary function test. Investigative Radiology 43(6), 395–402 (2008)

6. Sluimer, I.C., Prokop, M., Hartmann, I., van Ginneken, B.: Automated classification of hyperlucency, fibrosis, ground glass, solid, and focal lesions in high–resolution CT of the lung. Medical Physics 33(7), 2610–2620 (2006)
7. Depeursinge, A., Van De Ville, D., Platon, A., Geissbuhler, A., Poletti, P.-A., Müller, H.: Near-affine-invariant texture learning for lung tissue analysis using isotropic wavelet frames. IEEE Transactions on Information Technology in BioMedicine (2012)
8. Ojala, T., Pietikäinen, M., Mäenpää, T.: Multiresolution gray–scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(7), 971–987 (2002)
9. Haralick, R.M., Shanmugam, K., Dinstein, I.: Textural features for image classification. IEEE Transactions on Systems, Man and Cybernetics 3(6), 610–621 (1973)
10. Jain, A.K., Karu, K.: Learning texture discrimination masks. IEEE Transactions on Pattern Analysis and Machine Intelligence 18(2), 195–205 (1996)
11. Depeursinge, A., Foncubierta-Rodriguez, A., Van de Ville, D., Müller, H.: Lung Texture Classification Using Locally–Oriented Riesz Components. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part III. LNCS, vol. 6893, pp. 231–238. Springer, Heidelberg (2011)
12. Depeursinge, A., Vargas, A., Platon, A., Geissbuhler, A., Poletti, P.-A., Müller, H.: Building a reference multimedia database for interstitial lung diseases. Computerized Medical Imaging and Graphics 36(3), 227–238 (2012)
13. Unser, M., Van De Ville, D.: Wavelet steerability and the higher–order Riesz transform. IEEE Transactions on Image Processing 19(3), 636–652 (2010)
14. Unser, M., Van De Ville, D., Chenouard, N.: Steerable pyramids and tight wavelet frames in $L_2(\mathbb{R}^d)$. IEEE Transactions on Image Processing 20(10), 2705–2721 (2011)
15. Freeman, W.T., Adelson, E.H.: The design and use of steerable filters. IEEE Transactions on Pattern Analysis and Machine Intelligence 13(9), 891–906 (1991)
16. Greenspan, H., Belongie, S., Goodman, R., Perona, P., Rakshit, S., Anderson, C.H.: Overcomplete steerable pyramid filters and rotation invariance. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), pp. 222–228 (1994)
17. Vapnik, V.N.: The Nature of Statistical Learning Theory. Springer, New York (1995)

Blood Flow Simulation for the Liver after a Virtual Right Lobe Hepatectomy

Harvey Ho[1], Keagan Sorrell[2], Adam Bartlett[3], and Peter Hunter[1]

[1] Bioengineering Institute, University of Auckland, New Zealand
{harvey.ho,p.hunter}@auckland.ac.nz
[2] Dept. of Mechanical Engineering, University of Auckland, Auckland, New Zealand
[3] Dept. of Surgery, University of Auckland, Auckland, New Zealand

Abstract. In this paper we present a hybrid 0D-3D modeling method to investigate the hepatic flow in a virtual right lobe hepatectomy (RLH), the surgical procedure for adult-to-adult living donor liver transplanation (LDLT). The 3D method is employed to simulate complex 3D flow in the portal vein, and the 0D model is used to study the systemic hepatic circulation. In particular, we quantify the flow velocity and wall shear stress (WSS) in the left portal vein which increase dramatically post-RLH, and also simulate the essential hepatic distribution features in a healthy adult pre- and post-procedure. We further predict the arterial flow in the remnant left liver, which would decrease due to a hepatic arterial buffer response (HABR) effect. Finally we discuss the physiological significance of these phenomena, and the potential of this hybrid modeling approach.

1 Introduction

Liver transplantation is the treatment of choice for patients with end-stage liver disease [1]. However, there is a huge and growing disparity between supply and demand for cadaveric liver donors. Some patients on the waitlist are delisted due to deteriorated condition or death while waiting for a cadaveric liver to become available. One solution to alleviate this problem is the living donor liver transplantation (LDLT), whereby a portion of a living donor's liver is resected and transplanted to a recipient. Fig. 1 illustrates a LDLT scenario. Fig. 1(a) shows a whole liver and its vascular systems, which include a portal venous (PV) tree and a hepatic arterial (HA) tree suppling blood to the liver, and a hepatic venous (HV) tree which drains the blood into the inferior vena cava (IVC). In a right lobe hepatectomy (RLH), the larger right lobe is harvested and transplanted to an adult recipient. The remnant left lobe, about 30-40% of total liver mass, is perfused by the left portal vein (LPV), left hepatic artery (LHA) and drained by the left hepatic vein (LHV).

It has been reported that dramatic hepatic flow alterations would occur in both donor and recipient after the procedure [2,3]. For instance, in the left lobe of the donor, which previously only receives about 30-40% of total portal flow,

now receives the full portal inflow. Furthermore, the cardiac output after a lobectomy will increase [2]. This leads to an elevated total portal inflow immediately after the procedure. These dual effects would cause portal hyperperfusion, which actually is one of the major mechanisms driving liver regeneration following resection [3]. However, RLH carries with it the risk of donor mortality and mobidity. Therefore a good understanding of the perfusion pattern of the remnant liver is essential [2].
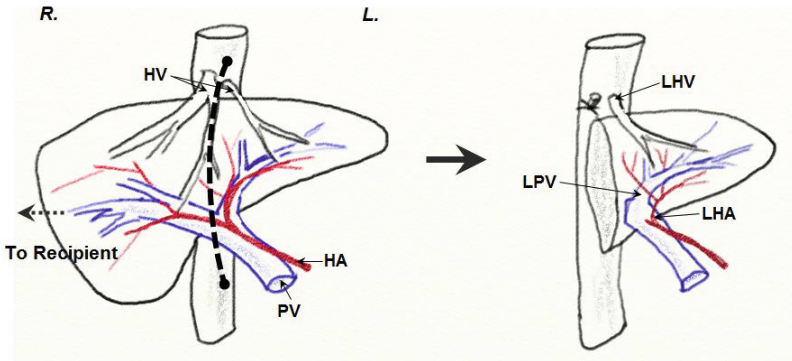
Fig. 1. Vascular anatomy of the liver: (a) the whole liver is supplied by a PV tree and a HA tree, and drained by a HV tree; (b) the remnant left liver, about 30-40% of total liver volume, is supplied by the l. HA and l. PV, and drained by l. HV. The bold broken curve represents the incision line. Abbreviations: HV - hepatic vein; HA - hepatic artery; PV - portal vein.

Three-dimensional (3D) flow in portal vein pre- and post-RLH was previously simulated by Ho *et al* [3]. They showed that increased portal flow and more complex flow streamlines occurred after the procedure, and speculated that this might be one of the main mechanisms that trigger a portal vein remodeling [3]. The systemic circulation pattern in the remnant liver, such as what would be the overall perfusion pattern, and how would arterial flow change with respect to increased PV flow, remain uninvestigated from a computational perspective.

In this work we propose a hybrid 0D-3D method to study hepatic flow from different perspectives, which could better prepare us to address these questions. We will reconstruct hepatic venous structures from a CT image, perform a virtual RLH, and simulate flow variations due to the virtual procedure. We will show the simulation results, and finally discuss the significance of this work.

2   Method

2.1   Medical Imaging and Vascular Construction

We studied the CT image (GE LightSpeed) of a male patient, who was admitted to the hospital due to a pathological condition (aneurysm) not relevant to the

liver. The spatial resolution of the image was $0.879 \times 0.879 \times 0.625$mm. The image, shown in Fig. 2(a), was used to visualize the portal and hepatic veins. Using a MIMICS software (Materialise, Leuven, Belgium) we segmented the liver and intrahepatic PV and HV trees. Also segmented are the superior mesenteric vein (SMV) and the splenic vein (SV), which merge into the PV (see Fig. 2b). To facilitate 3D flow simulation, the portal veins downstream the second generation were discarded because the image resolution was not high enough for us to conduct an accurate 3D vascular surface-reconstruction. A virtual incision line, indicated by the red line in Fig. 2(c), was assumed to occur at the right portal vein.
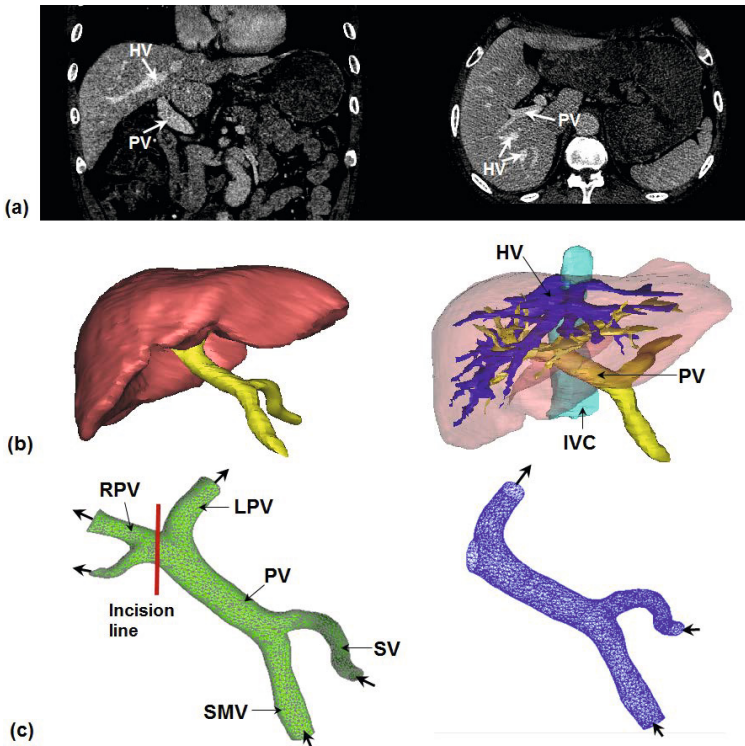
Fig. 2. Medical imaging and vascular construction: (a) front and axial view of the CT-image; (b) segmented PV and HV trees; (c) the portal vein is trimmed after the second generation. The red line indicates incision position and the bold arrows indicate flow directions.

The surface mesh of the PV, SMV and SV (see Fig. 2c) was imported into an ICEM software (ANSYS Inc.) for computational grid generation. The number of generated tetrahedral elements was about 160K.

2.2   Blood Flow Modeling

A hybrid 0D and 3D method was employed for blood flow modeling, whereby the 3D method was used for the simulation of complex 3D flow, and the 0D model for systemic circulation analysis, as briefed below.

3D Flow Simulation. The governing equations for blood flow in large arteries, the Navier Stokes equations, are expressed as:

$$\nabla \cdot \mathbf{v} = 0 \tag{1}$$

$$\rho(\frac{\partial \mathbf{v}}{\partial \mathbf{t}} + \mathbf{v} \cdot \nabla \mathbf{v}) = -\nabla p + \nabla \cdot \tau \tag{2}$$

where $\mathbf{v}$ and ρ represent the flow velocity and density of blood, respectively. p and τ represent the blood pressure and shear stress. The equations (1-2) were discretized over the computational grid of the portal system shown in Fig. 2(c), and were solved numerically using a finite volume based computational fluid dynamics solver, ANSYS CFX [4]. The portal flow was considered as steady, as revealed by ultrasonic measurements [3]. The inflow velocity boundary conditions, 20 cm/s and 30 cm/s, were prescribed at SV and SMV, respectively. These data were adopted from literature [2,3]. In addition, a zero pressure was imposed at the outlet(s) to allow free outflow.

0D Flow Simulation. For hepatic circulation models we follow the strategy of Debautt *et al* [5], who employed a dual source circuit to study hepatic perfusion. The difference between our model and that of [5] is that, instead of employing a large number of pi-filters for various PV, HV and HA generations, we simply used one pi-filter for each tree, therefore significantly simplified the circuit. The circuit, shown in Fig. 3, represents a basic hepatic circulation model. The direct current (DC) source in the circuit generates portal flow, whilst the pulsatile current (PC) source produces arterial flow. Note, that the law of mass conservation i.e. the hepatic inflow equals hepatic outflow ($F_{PV} + F_{HA} = F_{HV}$) is naturally obeyed in the circuit.

The parameters of the electronic components in a pi-filter (shown in Fig. 3b) are calculated based on a set of equations, e.g., [5]:

$$R_S = \frac{8\mu l}{\pi r^4} \tag{3}$$

$$L = \frac{1.33\rho l}{\pi r^2} \tag{4}$$

where μ is the viscosity of blood, l, r are the length and radius of blood vessel, respectively. The hepatic arterial resistance R_{HA} is a nonlinear element to simulate a hepatic arterial buffer response (HABR), the intrinsic regulating mechanism whereby the changes of hepatic arterial flow counteracts that of portal flow [6].
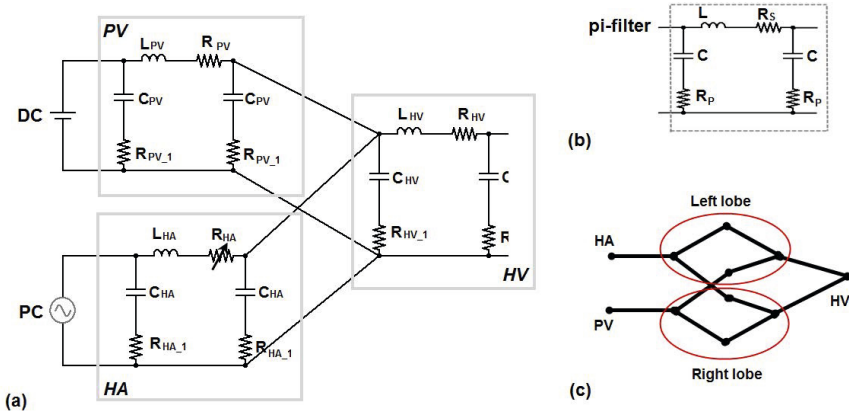
Fig. 3. Electrical analog circuits for hepatic circulation. (a) Basic model: the three pi-filters simulate flow in HA, HV and PV trees (each shown inside a gray rectangle); (b) a pi-filter; (c) the schema of the extended circuit accounts for the hepatic circulation to the left and right lobes. The parameter values of the electric components are shown in Table 1.

Table 1 lists the parameters for the electronic components in Fig. 3(a). Their values are based on the data in [5], and were fine-tuned after simulations.

Since the left and right lobes of the whole liver are perfused and drained independently, the basic electric circuit of Fig. 3(a) was expanded into a more complex circuit that differentiates the left and right lobes. The schema of this electric circuit is shown in Fig. 3(c). Note, that each segment in the circuit represents a pi-filter, and that the two lobes receive different portion (left: 40%, right: 60%) of total HA and PV flow, with each lobe drained by a single hepatic vein.

3   Results

3.1   3D Flow Alteration in the Portal Vein

The 3D flow simulation took about 30 minutes to complete on a desktop computer (Intel Core 2.4GHz). A variety of hemodynamic data were computed. Of particular interest are the flow velocity and the wall shear stress (WSS) pre- and post the virtual RLH procedure.

Fig. 4(a) visualizes flow streamlines along the flow path. It shows that the peak flow velocity in the LPV increased from 18cm/s to 46cm/s, as a result of receiving full portal flow. Consequently, the WSS in the LPV, shown in Fig. 4(b), was drastically elevated from merely 0.2Pa pre-RLH to 0.6Pa post-RLH. This agrees with the simulation of [3] which showed that WSS almost doubled post-RLH (from 0.4 Pa to 0.8 Pa). Also observable are the strong helical flows (indicated by slim arrows) in PV developed after the merging point of the SMV and SV.

Table 1. Parameters of Electronic components in Fig. 3(a)

Component	Whole organ	Unit
R_{HA}	1.5689e9 (nominal value)	$Pa \cdot s/m^3$
R_{HV}	1.3630e8	$Pa \cdot s/m^3$
R_{PV}	3.7807e7	$Pa \cdot s/m^3$
R_{HA_1}	1.562e10	$Pa \cdot s/m^3$
R_{HV_1}	1.40607e7	$Pa \cdot s/m^3$
R_{PV_1}	2.1054e8	$Pa \cdot s/m^3$
R_{IVC}	16.5e6	$Pa \cdot s/m^3$
C_{HA}	1.2803e-10	$m^3/Pa \cdot s$
C_{HV}	1.4224e-6	$m^3/Pa \cdot s$
C_{PV}	9.4995e-9	$m^3/Pa \cdot s$
L_{HA}	200e6	$Pa \cdot s^2/m^3$
L_{HV}	30e6	$Pa \cdot s^2/m^3$
L_{PV}	50e6	$Pa \cdot s^2/m^3$

This was not presented in [3] because that study placed the flow inlet after the SMV-SV junction. The physiological implications of the helical flows to PV wall remodeling remain to be investigated.

3.2   Hepatic Circulation in the Whole Liver and the Left Lobe

The portal vein pressure was assumed to be 6 mmHg at DC, and a pulsatile pressure varying between 80mmHg and 120mmHg was prescribed from the PC. The resulted flow rate waveforms of PV, HA and HV in the whole liver are plotted in Fig. 5(a). In particular, the total hepatic flow (F_{HV}) is about 1.45 L/min, i.e. approximately 30% of the cardiac output ($\sim$5 L/min) in a healthy adult. The hepatic arterial flow, which is more pulsatile than the portal and hepatic venous flows, contributes about 400mL/min (or 33%$\approx$1/3 of total hepatic flow volume) oxygenated blood to hepatic circulation, whilst the portal vein supplies the remaining 1,050 mL/min (or 67%$\approx$ 2/3 of total volume) poorly-oxygenated yet nutrient-borne blood.

The above flow distribution represents the essential hepatic circulation features in a healthy adult, and was achieved by using the electric circuit of Fig. 3(a). We further analyzed the flow distribution into the left and right lobes using the extended circuit of Fig. 3(c). Fig. 5(b) shows the results: the flow rate in LPV is 1.05 L/min post-RLH, almost threefold of that pre-RLH (380 mL/min). This causes portal hyperperfusion (and high shear) in the LPV, as also revealed from 3D simulations. On the other hand, the left arterial flow decreases from 140 mL/min pre-RLH to 50 mL/min post-RLH. This surprising phenomenon is due to the HABR that was observed in partial liver grafts in recipients [7]. The presumed aim of this intrinsic mechanism is to keep the total hepatic circulation within a physiological range.
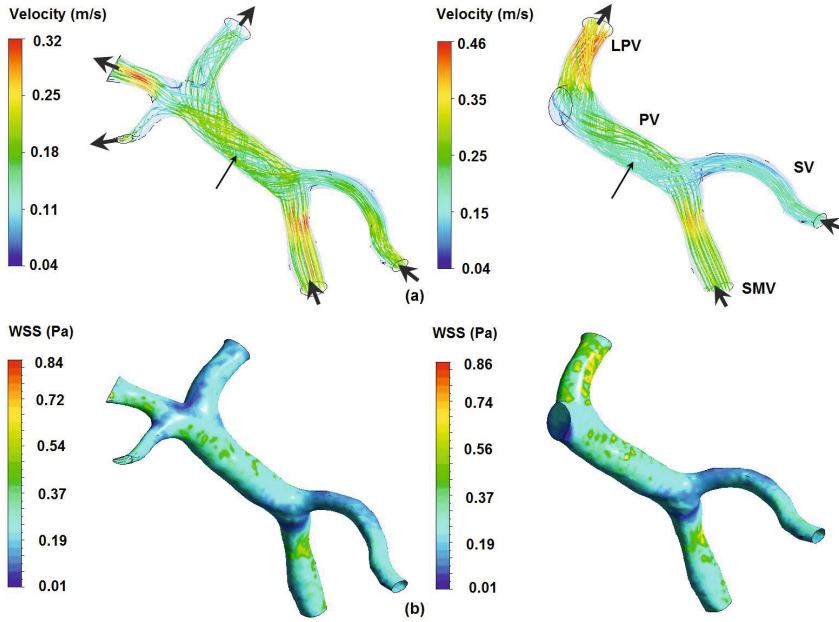
Fig. 4. Comparison of hemodynamic data pre- and post-RLH (a) flow streamline: helical flows are formed after the SMV and SV junction. The flow velocity increase in the LPV from 18cm/s to 46cm/s; (b) wall shear stress in the LPV is drastically increased, from merely 0.2Pa to 0.6Pa.

4 Discussion

Right lobe hepatectomy is the surgical method used in adult-to-adult LDLT [1]. Donor mortality and morbidity may occur due to multiple factors such as inadequate remnant liver mass, compromised hepatic circulation, etc. A comprehensive understanding of hepatic circulation in the remnant liver would aid biomedical and surgical research of this procedure. In this paper we used a hybrid 3D-0D method for the flow analysis of a virtual RLH. In particular, we simulated the essential flow distribution feature in the whole liver, and also in the remaining left lobe. This has never been performed previously, to our knowledge. Moreover, we made a physiological predication that hepatic arterial flow may decrease sharply in the liver remnant following RLH. This decrease, after verified through *in vivo* ultrasonic measurements, could be of significance. The possible implications of such a decrease are many, for example hepatic arterial thrombosis is a life-threatening disorder that is associated with liver dysfunction.

The presented models can be further extended to study more complex and clinically-related hepatic circulation problems. For instance, the pathological condition of the liver, e.g., cirrhosis, may be taken into the circulation model by altering the hepatic venous resistance and raising the portal pressure.
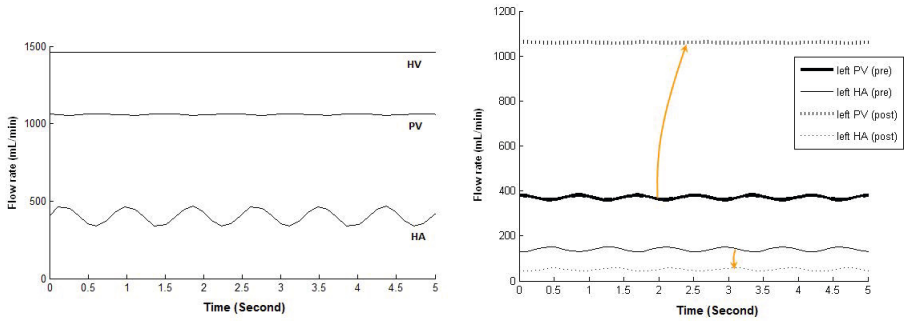
Fig. 5. (a) Flow rate in the HV, HA and PV in a whole liver: HA contributes 1/3 of total hepatic flow whilst PV contributes the rest; (b) LPV flow drastically increases but the LHA flow decreases due to the HABR effect

5   Conclusion

In this paper we simulated blood flow in a virtual right lobe hepactomy that was constructed from a 3D CT image. We showed the increased flow velocity and wall shear stress in portal vein, and reproduced the essential flow distribution features in the whole liver and the remnant left lobe.

References

1. Lo, C.M., Fan, S.T., Liu, C.L., Wei, W.I., Lo, R.J., Lai, C.L., Chan, J.K., Ng, I.O., Fung, A., Wong, J.: Adult-to-adult living donor liver transplantation using extended right lobe grafts. Annals of Surgery 226(3), 261–270 (1997)
2. Niemann, C.U., Roberts, J.P., Ascher, N.L., Yost, C.S.: Intraoperative hemodynamics and liver function in adulttoadult living liver donors. Liver Transplantation 8(12), 1126–1132 (2002)
3. Ho, C., Lin, R., Tsai, S., Hu, R., Liang, P., Sheu, T.W., Lee, P.: Simulation of portal hemodynamic changes in a donor after right hepatectomy. Journal of Biomechanical Engineering 132(4), 041002–041007 (2010)
4. ANSYS: ANSYS CFX-Solver, Release 10.0: Theory. ANSYS Europe Ltd. (2005)
5. Debbaut, C., Monbaliu, D., Casteleyn, C., Cornillie, P., Loo, D.V., Masschaele, B., Pirenne, J., Simoens, P., Hoorebeke, L.V., Segers, P.: From vascular corrosion cast to electrical analog model for the study of human liver hemodynamics and perfusion. IEEE Transactions on Biomedical Engineering 58(1), 25–35 (2011)
6. Lautt, W.W., Greenway, C.V.: Conceptual review of the hepatic vascular bed. Hepatology 7(5), 952–963 (1987)
7. Marcos, A., Olzinski, A., Ham, J., Fisher, R., Posner, M.: The interrelationship between portal and arterial blood flow after adult to adult living donor liver transplantation. Transplantation 70(12), 1697–1703 (2000)

A Combinatorial Method for 3D Landmark-Based Morphometry: Application to the Study of Coronal Craniosynostosis[⋆]

Emeric Gioan, Kevin Sol, and Gérard Subsol

LIRMM, CNRS / Université Montpellier 2, France

Abstract. We present a new method to analyze, classify and characterize 3D landmark-based shapes. It is based on a framework provided by oriented matroid theory, that is on a combinatorial encoding of convexity properties. We apply this method to a set of skull shapes presenting various types of coronal craniosynostosis.

1 Introduction

Since three decades, *Geometry Morphometrics* has revolutionized the quantitative analysis of the variation of the shape of anatomical structures [1]. It can be defined as a collection of methods that process directly the coordinates of *landmarks*, in 2D or in 3D, rather than with traditional distance or angle measurements. The landmarks, which are in general points, can be defined by experts in anatomy or can be automatically computed by geometrical feature extraction algorithms. Landmark-based morphometry methods are now used in many medical applications but they present some important drawbacks as emphasised in [7,6].

In the *superposition methods*, many schemes of alignment have been proposed to superimpose the sets of landmarks (e.g. Procrustes, matching of a specified edge, etc.). The morphometrical result is then directly related to the alignment scheme itself, which is often selected by guessing the potential variability of the shape. Moreover, in some cases, the alignment scheme may result in transforming a strictly local deformation into a global one (*Pinocchio effect* [4]).

In the *deformation methods*, the difficulty is to define an appropriate class of transformations to warp a set of landmarks toward another. If the class is too general, the reference shape may deform to anything without any geometrical consistency. On the contrary, if the transformation is too constrained, there may be no accurate registration. In both cases, the transformation parameters will not properly characterize the shape deformation. Many mathematical formulations have been proposed but they remain difficult to assess and interpret as we do not know the potential variability of the shape.

Linear distance-based methods as the Euclidean Distance Matrix Analysis do not require to define any geometrical transformation to align or deform the shape. Nevertheless, the results which are based on inter-landmark distances are quite

---

difficult to understand. The practitioner, indeed, does not only want not to know that the shape of an anatomical structure has varied by a quantity, but also to graduate this quantity on a standard scale and to localize the deformation. And even if some procedures to identify the most influential landmarks were proposed [3], this kind of methods remains little used for medical applications.

In this paper, we propose to encode 3D landmark configurations as oriented matroids, a combinatorial mathematical structure which was developed over the past forty years [2]. The idea is to code all the relative 3D positions of points of the shape, without taking into account the distances between them. As in linear distance-based methods, no alignment or deformation is to be computed. Using tetrahedra orientations in the shape, we obtain a vector of discrete values (-1 or 1) which characterizes the structure of the shape, independently of its size, its position, or more generally of any linear isomorphism of the space (see Section 2). The mathematical structure allows one to detect structural changes in a shape such as the crossing of a landmark through the plane defined by three others, and more generally all convexity properties involving subsets of landmarks. In Section 3, we introduce some theoretical and computational method to analyze and compare those shape representations. Let us point out that the discretization allows to perform a morphometrical analysis, without any numerical error or approximation. We show in Section 4 a clinical application to the classification of coronal craniosynostosis, yielding simple formal/geometrical/combinatorial characterizations of the classes within the studied set of individuals.

2   Combinatorial Structure of a 3D Model

Let E be a finite set of n labels. Let M be a set of n points labelled by E in the real affine space of dimension 3. We call M a *model*. We assume that these n points are in general position, meaning that every subset of E having four elements is a basis of the affine space. We call *basis* such a subset $B = \{a, b, c, d\}$. Let $\mathcal{B}$ be the set of all bases. For the sake of formal simplicity, we assume that E and $\mathcal{B}$ are linearly ordered. Hence every element $B = \{a, b, c, d\}_<$ of $\mathcal{B}$ is ordered, and we denote $B = abcd$. Then, for instance, $\mathcal{B}$ can be ordered lexicographically.

Fig. 1. The two possible orientations of tetrahedron $abcd$ given by $\chi_M(abcd)$

Every basis B in $\mathcal{B}$ gets a sign in $\{+1, -1\}$, called *orientation* or *sign* of B in M, and denoted $\chi_M(B)$. It is defined as the orientation of the tetrahedron formed by the four points in B with respect to the ordering of B and a chosen orientation of the space. See Figure 1.

Equivalently, in linear algebra terms, and in computational terms, $\chi_M(B)$ is the sign of the determinant of the 4×4 matrix whose columns give the coordinates of the points in a canonical basis of the space, that is:

$$\chi_M(abcd) = \text{sign}\left(\text{determinant} \begin{pmatrix} 1 & 1 & 1 & 1 \\ x_a & x_b & x_c & x_d \\ y_a & y_b & y_c & y_d \\ z_a & z_b & z_c & z_d \end{pmatrix} \right).$$

Let χ_M be the list of signs of elements of $\mathcal{B}$ in M, with respect to the linear ordering of $\mathcal{B}$, that is $\chi_M = [\chi_M(B_1), \chi_M(B_2), \chi_M(B_3)\ldots]$. We call χ_M the *chirotope* of M. We consider χ_M as a vector in the real space $\mathbb{R}^{\mathcal{B}}$ of dimension $|\mathcal{B}|$; and for a vector x of this space, $x(B)$ denotes the value of x at coordinate B.

Figure 2 shows an example of M on 5 points labelled by $1 < 2 < ... < 5$. The following table shows the ordered list of bases and their signs:

basis	1234	1235	1245	1345	2345
sign	-1	-1	$+1$	-1	-1

yielding the chirotope $\chi_M = [-1, -1, +1, -1, -1] \in \mathbb{R}^5$.

Figure 3 shows another example M' on 5 points with the same labels. The chirotope of M' is $\chi_{M'} = [+1, -1, +1, -1, -1]$. In comparison with M from Figure 2, the point 1 has crossed the plane 234 (and no other plane spanned by points of the model). That is why χ_M and χ'_M differ only with the sign of the basis 1234. Also, for example, it can be read from the chirotope if the point 1 belongs or not to the convex hull formed by the other points.

Observe also that if a point moves in M without crossing a plane spanned by other points, then χ_M does not change. The properties encoded by χ_M do not depend directly on numerical measures (distances, angles...), and this combinatorial encoding rather describes the "structural shape" of the point set.

The properties of χ_M are known as the *uniform oriented matroid theory* [2]. This rich mathematical theory is forty years old and can be seen as a combinatorial abstraction of linear algebra. In the case of real points, the oriented matroid defined by χ_M is a mathematical structure which is combinatorial (i.e. defined on a finite set), encodes the relative positions of the points of M, and catches in particular all their convexity properties[1].

In this paper, we use this information to compare models, labelled by a same set, practically given by landmarks. Formally, we fix a set $\mathcal{M}$ of models M labelled by E, and we study the set of χ_M, for $M \in \mathcal{M}$.
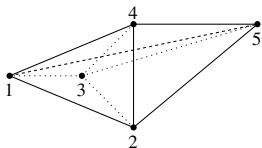
Fig. 2. A model M. Full lines belong to the front part of the convex hull; light dashed lines belong to its back part; the medium dashed line 15 is inside the convex hull.
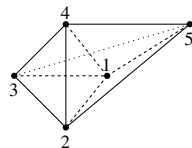
Fig. 3. A model M' obtained from M by moving 1 inside the tetrahedron 2345. The lines 12, 13, 14 and 15 are now inside the convex hull.

---

[1] We mention that the case where points may not be in general position leads to a possible 0 sign for the values $\chi_M(B)$, and to the general theory of oriented matroids. In this paper, for the sake of clarity and concision, we focus on the uniform case.

3   Method of Analysis

3.1   General Scope

The main concept to compare models in $\mathcal{M}$ is the following. For a subset $\mathcal{D}$ of $\mathcal{B}$, and two vectors x and y in $\mathbb{R}^{\mathcal{B}}$, we define the *distance between x and y w.r.t. $\mathcal{D}$* as

$$d_{\mathcal{D}}(x,y) = (1/2) \cdot \sum_{B \in \mathcal{D}} \mid x(B) - y(B) \mid$$

which is, mathematically, a distance for projections onto the space $\mathbb{R}^{\mathcal{D}}$. In what follows we make the abuse to call $d_{\mathcal{D}}$ a distance in $\mathbb{R}^{\mathcal{B}}$ even when $\mathcal{D} \neq \mathcal{B}$. Observe that the distance $d_{\mathcal{D}}$ between two chirotopes χ_M and $\chi_{M'}$ equals the number of bases in $\mathcal{D}$ having different signs in M and M'.

Let us now consider a partition of $\mathcal{M}$ in two classes $\mathcal{M} = \mathcal{C} \uplus \mathcal{C}'$, or, more generally, a partition into k classes $\mathcal{M} = \mathcal{C}_1 \uplus \ldots \uplus \mathcal{C}_k$. Such a partition can be given by the experts that provided the models (e.g. four known medical types for skulls as in Section 4). On the other hand, such a partition can be built using the chirotopes of the models without any other preliminary knowledge. Note that if classes satisfying automatic unsupervised classification criteria match classes given by experts, then it means that the combinatorial data captures some geometrical properties that characterize the classes in which the experts are interested, and yields a formal characterization of those classes.

Once a partition is given, either by experts or by automatic classification, a second step is to characterize classes using the less possible information from the chirotopes. For instance, we look for a few bases whose signs allow to determine if a model belongs to a given class or not. In terms of applications, those bases should be significant, providing either a mathematical formal confirmation of criteria used by experts, or new properties pointing out typical features of the class that would interest experts.

3.2   Automatic Classification

Given a set $\mathcal{C}$ of models, the *barycenter $m_{\mathcal{C}}$ of $\mathcal{C}$* is the vector in $\mathbb{R}^{\mathcal{B}}$ whose coordinate at basis B is the mean of the signs of models in $\mathcal{C}$ at basis B, that is:

$$m_{\mathcal{C}}(B) = \tfrac{1}{|\mathcal{C}|} \sum_{M \in \mathcal{C}} \chi_M(B).$$

We say that the partition $\mathcal{M} = \mathcal{C} \uplus \mathcal{C}'$ satisfies the *k-means criterion* if for every $M \in \mathcal{M}$ we have: $M \in \mathcal{C}$ if and only if $d_{\mathcal{B}}(\chi_M, m_{\mathcal{C}}) < d_{\mathcal{B}}(\chi_M, m_{\mathcal{C}'})$, and $M \in \mathcal{C}'$ if and only if $d_{\mathcal{B}}(\chi_M, m_{\mathcal{C}}) > d_{\mathcal{B}}(\chi_M, m_{\mathcal{C}'})$. Of course, this criterion directly extends to a partition into k classes.

Various clustering algorithms can be used to build automatically a partition into classes satisfying criteria such as the above one. The most classical one consists in initializing arbitrary classes, computing the barycenters for each class and move a model to an other class if it is closer to the barycenter of this other class than to the barycenter of its initial class. Repeating this procedure until

stability yields a partition satisfying the k-means criterion. Then, one may test many initializations and select an "optimal" classification.

For lack of space, we just mention that such an algorithm allows to detect "mean" individuals among the classes, characterized by "mean" chirotopes, and that other clusterisation algorithms can be used (e.g. k-medoids).

3.3   Characterization of Classes

Assume that a partition $\mathcal{M} = \mathcal{C} \uplus \mathcal{C}'$ is given. We now look for simple and significant combinatorial criteria defining this partition.

Formally, in this paper, we will build classes depending on three parameters: a set $\mathcal{D}$ of bases, a vector x in the space $\mathbb{R}^{\mathcal{B}}$, and a real value l. We define

$$\mathcal{C}_{\mathcal{D},x,l} = \{M \in \mathcal{M} \mid d_{\mathcal{D}}(\chi_M, x) \leq l\}.$$

This set can be considered as the set of model chirotopes contained in a "ball" in the space $\mathbb{R}^{\mathcal{B}}$, centered at x and of radius l for the distance $d_{\mathcal{D}}$.

Practically, we look for a subset $\mathcal{D}$ of $\mathcal{B}$ the smallest possible for simplicity, for a vector x in $\mathbb{R}^{\mathcal{D}}$ with coordinates in $\{-1, +1\}$, and for an integer value l, such that $\mathcal{C}_{\mathcal{D},x,l}$ is equal to class $\mathcal{C}$. In this combinatorial setting, a model M belongs to the class $\mathcal{C}$ if and only if there are at most l bases B in $\mathcal{D}$ such that $\chi_M(B) \neq x(B)$. Hence, $\mathcal{D}$ serves as a significant set of bases, x as the characteristic values of these bases for $\mathcal{C}$, and l as a threshold value to the fact that a model fits this charateristic values.

As an example of particular interest, consider the case where $\mathcal{D}$ consists of a single basis $\mathcal{D} = \{B\}$. Then x is given by a single sign, say $x(B) = +1$. Then a model M belongs to $\mathcal{C}$ if and only if $\chi_M(B) = +1$, and it belongs to $\mathcal{C}'$ if and only if $\chi_M(B) = -1$. Equivalently, we have $\mathcal{C} = \mathcal{C}_{\{B\},+1,0}$. In this case, the sign of B determines if M belongs to $\mathcal{C}$ or not. We say that B is *totally discriminant* for the class $\mathcal{C}$. This is the most simple possible characterization of a class.

The problem is to detect such parameters $\mathcal{D}, x, l$. Let us use the following procedure. For a basis B, we define the *discriminability* of B as $\tau(B, \mathcal{C}, \mathcal{C}') = \mid m_{\mathcal{C}}(B) - m_{\mathcal{C}'}(B) \mid /2$. This value belongs to $[0, 1]$. The closer to 1 this value is, the more significant the basis B is for the class $\mathcal{C}$. The extreme case is $\tau(B, \mathcal{C}, \mathcal{C}') = 1$, implying that $m_{\mathcal{C}}(B) = +1$ and $m_{\mathcal{C}'}(B) = -1$ (or the inverse), and then, by definition of $m_{\mathcal{C}}$ and $m_{\mathcal{C}'}$, that every model in $\mathcal{C}$ (resp. $\mathcal{C}'$) has value $+1$ (resp. -1), meaning that B is totally discriminant for $\mathcal{C}$ and $\mathcal{C}'$. On the contrary, $\tau(B, \mathcal{C}, \mathcal{C}') = 0$ can be obtained for instance when B has the same sign in every model in $\mathcal{M}$, or also at the limits when B has a random sign; hence we cannot expect B to be significant to describe $\mathcal{C}$.

We can order the bases in $\mathcal{B}$ according to their discriminability. This ordering allows a computation to search parameters $\mathcal{D}, x, l$. Indeed, from the computational viewpoint, there may be a huge number of elements in $\mathcal{B}$ (any set of four landmarks is a basis), and hence a non-affordable number of subsets $\mathcal{D}$ to test. Considering only bases with high discriminability allows to restrict the number of subsets to be tested, while focusing on the information that has some chance to be significant. So we consider bases having the highest possible

discriminability to build a set $\mathcal{D}$ and an integer l as required. The values $x(B)$ determining x are naturally given by the signs of $m_{\mathcal{C}}(B) - m_{\mathcal{C}'}(B)$ for $B \in \mathcal{D}$.

Observe that if a partition $\mathcal{M} = \mathcal{C}_1 \uplus \ldots \uplus \mathcal{C}_k$ is given and such parameters are known for every class $\mathcal{C}_i$ in the partition, then we have two independent criteria to detect a class: first the direct criterion provided by the parameters characterizing $\mathcal{C}_i$, secondly the negation of the criteria determining that a vector belongs to another class $\mathcal{C}_j$, $j \neq i$.

4 Applications to Coronal Craniosynostosis

Coronal craniosynostosis is a rare infant pathology in which one of the two coronal sutures of the skull prematurely fuses by ossification. We can distinguish between left (LUCS), resp. right (RUCS), uni-coronal craniosynostosis when only one side is affected and the bilateral case (BCS) when both sides are fused. It is essential to use 3D morphometric tools to quantify and analyze precisely the deformation of the skull shape in this pathology. In [5], the authors used classical landmark-based tools (Procrustes and Principal Component Analysis) to study a database of the International Craniosynostosys Consortium and assess the (a)symmetry of the pathology. We could test our new morphometric method on the same data and we propose below some preliminary results.

The clinical database is composed of CT-Scan images of 40 children diagnosed with non syndromic coronal craniosynsostis (LUCS=8 / RUCS=17 / BCS=15) and 20 unaffected individuals of the same age. Anatomy experts pointed 41 landmarks on these CT-Scan and added 92 semilandmarks along predefined curves, as illustrated on Figure 4. No alignment or spatial normalization was performed.

To compute the $\binom{133}{4}$ tetrahedron signs, it takes about 6 hours on a standard PC. This process is performed only once. The k-means clustering itself takes about 30 seconds on a standard PC, but one wants to run it multiple times (1000 in our application) with different starting positions.

First, we present results obtained by using the whole set of 133 landmarks. We started off by testing if the four classes given by the experts (BCS, LUCS, RUCS and Unaffected) satisfie the k-means criterion. This criterion is satisfied by these four classes, so we can assume that the premature fusion of sutures
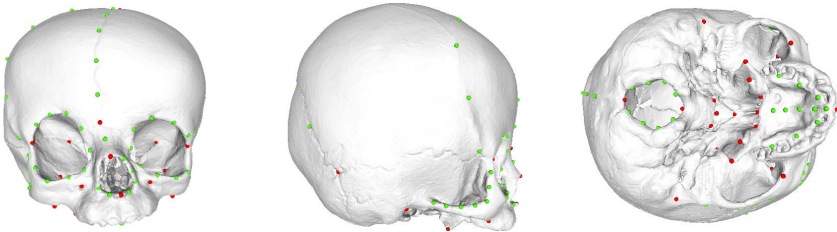
Fig. 4. Illustration of the 133 3D landmarks (RUCS skull shown here). Anatomical landmarks are in red and curve semilandmarks are displayed in green.

corresponds to changes in the skull shape chirotopes. We applied the k-means algorithm to find a partition of the 60 individuals in 4 clusters which minimizes the sum of distances between the individuals and the barycenter of their class. The algorithm returns exactly the classes defined by the experts, strengthening the above observation. With this result we obtain also for every class $\mathcal{C}$ the closest individual of the barycenter of $\mathcal{C}$.

Let $\mathcal{C}$ be a class among BCS, LUCS, RUCS and Unaffected. We denote by $\mathcal{C}'$ the set of all the individuals which are not in $\mathcal{C}$. We computed the discriminability $\tau(B, \mathcal{C}, \mathcal{C}')$ for every basis B and class $\mathcal{C}$. For every class there exist several bases which are totally discriminant. Precisely there are 147 such bases for BCS, 15,667 such bases for LUCS, 5,064 such bases for RUCS, and 7 such bases for Unaffected. So for each class we have many simple characterizations (direct criterion and negation of the criterion of the others classes).

In a second time we used only the 41 anatomical landmarks. The first reason is to focus only on what has a real anatomical meaning and a precise anatomical definition. The second reason is to speed up all the calculus as it reduces the number of tetrahedra from $\binom{133}{4} = 12,457,445$ to $\binom{41}{4} = 101,270$.

The k-means algorithm yields a partition into 4 clusters which match exactly the 4 classes given by the experts, except for one individual among the 60.

There exist bases which are totally discriminant for LUCS (22 bases) and RUCS (4 bases). The four bases for RUCS all contain the *bregma* (intersection of the coronal and the sagittal suture) and the *lambda* (intersection of the sagittal and lambdoidal sutures). Each basis contains a third landmark in the median sagittal plane. This third landmark is the *nasion* for two bases and the *nasale* for the two other bases. For each basis, the fourth landmark is a point of the right part of the cranial base. For each of these bases, if we replace the fourth landmark by the symmetric landmark with respect to the median sagittal plane, we obtain one of the 22 bases which are totally discriminant for LUCS. Figure 5 shows one of the four basis totally discriminant for RUCS and the symmetric basis which is totally discriminant for LUCS.

For BCS, there is no totally discriminant basis. Let us detail one characterization we obtained, among other ones. We found two tetrahedra B_1 and B_2 such that an individual is BCS if and only if both the orientations of these tetrahedra are -1. That is BCS $= \mathcal{C}_{\{B_1, B_2\}, x, 0}$ for $x(B_1) = x(B_2) = -1$. Their discriminabilities are high and equal 0.89 and 0.91. These two tetrahedra share two landmarks. Let us denote $B_1 = 1234$ and $B_2 = 1256$. Figure 6 shows these two tetrahedra on some BCS model, and Figure 7 details their positions for the BCS class. Figure 8 shows the other possibilities, available for models in the other classes. The involved landmarks are: the left *anterior clinoid process* (1); the left *asterion* (posterior end of the parietomastoid suture) (2); the left and the right *fronto-zygomatic junction* at orbital rim (4 and 3 respectively); the anterior *nasal spine* (5); and the left external *auditory meatus* (6).

For the unaffected individuals, we found (for instance) a slightly more involved characterization: a set $\mathcal{D}_2$, with five bases having discriminabilities between 0.85 and 0.875, such that this class equals $\mathcal{C}_{\mathcal{D}_2, y, 2}$ for some $y \in \{-1, +1\}^{\mathcal{B}}$. That is:
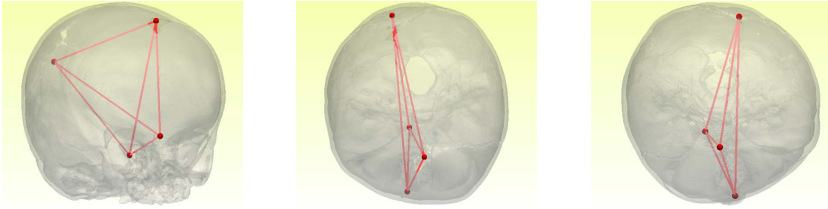
Fig. 5. One of the 4 bases totally discriminant for the RUCS (left and middle) and the basis obtained by symmetry which is totally discriminant for the LUCS (right)
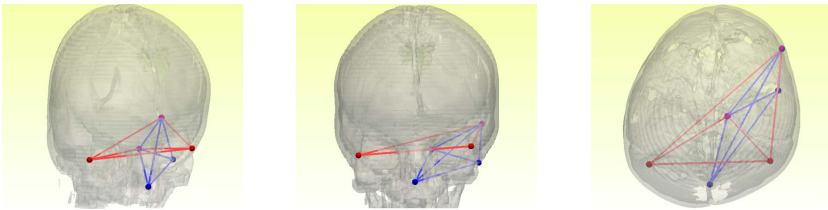
Fig. 6. The two tetrahedra (red and blue) which give a characterization for BCS
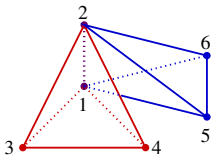
Fig. 7. The orientations $[-1, -1]$ of B_1 (red) and B_2 (blue) characterize BCS
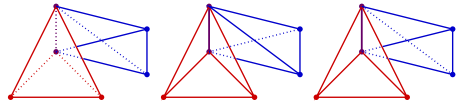
Fig. 8. The other possible orientations $[-1, +1]$, $[+1, -1]$, and $[+1, +1]$ (from left to right), available for non-BCS models.

x is unaffected if and only if there exist at least three bases in $\mathcal{D}_2$ such that the signs of these bases are the same in x and y.

Finally, we point out that the formal/geometrical/combinatorial characterizations above are obtained with respect to a given whole set of models, and may depend noticeably on this initial data. However, they raise the question of medical anatomical interpretations.

References

1. Adams, D.C., Rohlf, F.J., Slice, D.E.: Geometric Morphometrics: Ten Years of Progress Following the "Revolution". Ital. J. Zool. 71(1), 5–16 (2004)
2. Björner, A., Vergnas, M.L., Sturmfels, B., White, N., Ziegler, G.M.: Oriented Matroids, 2nd edn. Cambridge University Press (2000)

3. Cole III, T.M., Richtsmeier, J.T.: A Simple Method for Visualization of Influential Landmarks When Using Euclidean Distance Matrix Analysis. Am. J. of Phys. Anthropol. 107(3), 273–283 (1998)
4. von Cramon-Taubadel, N., Frazier, B.C., Lahr, M.M.: The Problem of Assessing Landmark Error in Geometric Morphometrics: Theory, Methods, and Modifications. Am. J. of Phys. Anthropol. 134(1), 24–35 (2007)
5. Heuzé, Y., Martínez-Abadías, N., Stella, J.M., Senders, C.W., Boyadjiev, S.A., Lo, L., Richtsmeier, J.T.: Unilateral and Bilateral Expression of a Quantitative Trait: Asymmetry and Symmetry in Coronal Craniosynostosis. J. Exp. Zool. Part B 318(2), 109–122 (2012)
6. Niu, J., Li, Z., Salvendy, G.: Mathematical Methods for Shape Analysis and form Comparison in 3D Anthropometry: A Literature Review. In: Duffy, V.G. (ed.) Digital Human Modeling, HCII 2007. LNCS, vol. 4561, pp. 161–170. Springer, Heidelberg (2007)
7. Richtsmeier, J.T., Deleon, V.B., Lele, S.R.: The Promise of Geometric Morphometrics. Yearbook of Physcial Anthropology 45, 63–91 (2002)

A Comprehensive Framework for the Detection of Individual Brain Perfusion Abnormalities Using Arterial Spin Labeling

Camille Maumet[1,2,3,4], Pierre Maurel[1,2,3,4], Jean-Christophe Ferré[1,2,3,4,5], and Christian Barillot[1,2,3,4]

[1] University of Rennes 1, Faculty of Medecine, F-35043 Rennes, France
[2] INSERM, U746, F-35042 Rennes, France
[3] CNRS, IRISA, UMR 6074, F-35042 Rennes, France
[4] Inria, VISAGES Project-Team, F-35042 Rennes, France
[5] CHU Rennes, Department of Neuroradiology, F-35033 Rennes, France

Abstract. Arterial Spin Labeling (ASL) enables measuring cerebral blood flow in MRI without injection of a contrast agent. Perfusion measured by ASL carries relevant information for patients suffering from pathologies associated with singular perfusion patterns. However, to date, individual identification of abnormal perfusion patterns in ASL usually relies on visual inspection or manual delineation of regions of interest.

In this paper, we introduce a new framework to automatically outline patterns of abnormal perfusion in individual patients by means of an ASL template. We compare two models of normal perfusion and assess the quality of detections comparing an *a contrario* approach to the Generalized Linear Model (GLM).

1 Introduction

Perfusion is the process through which the blood provides nutrients and oxygen to the tissues by means of micro-circulation. ASL is a recent MRI technique [1] that allows perfusion measurement and quantification of Cerebral Blood Flow (CBF). Contrary to Dynamic Susceptibility weighted Contrast (DSC) imaging – the most validated technique to measure perfusion with MRI – ASL does not rely on the injection of an exogenous contrast agent. In a few words, blood water is labeled with a radio-frequency pulse in the neck and, after a delay called inversion time, a labeled image of the brain is acquired. The difference between the label image and a control image, acquired without labeling, leads to a perfusion weighted image. Due to the low signal to noise ratio (SNR) of the ASL sequence, a single pair of control and label image is not sufficient to measure perfusion, the acquisition is usually repeated several times leading to R pairs of images. Perfusion information is then extracted by averaging. A model is applied to this image to obtain a quantification of CBF. ASL is particularly well suited for longitudinal studies or studies of patients with difficult venous access such as children. Its non-invasiveness makes ASL the method of choice

for the study of perfusion of healthy subjects. However this comes at the cost of low SNR and lower spatial resolution than DSC.

Today, ASL studies are able to outline differential patterns of perfusion in pathological populations, including Alzheimer disease [2] or schizophrenic patients. At the same time, the interest of ASL in the study of individual patients presenting singular perfusion abnormalities was also demonstrated. In these pathologies, the highly variable pathological patterns preclude group analysis. Abnormal patterns of perfusion must then be identified at the subject level. It is for instance the case of tumors, strokes or multiple sclerosis lesions. For patients diagnosed with tumors, the clinician is interested by hyperperfusions that would reveal the grade of the tumor [3]. In this context, perfusion abnormality studies usually rely on comparing the level of perfusion in the lesion to the controlateral normal tissue. This method is based on manual regions of interest delineations, a time-consuming task prone to inter-expert variability.

The aim of this paper is to present an automatic framework to identify hypoperfused and hyperperfused regions in individual patients by comparison to a model of normal perfusion. In [4], we investigated the ability to detect hypoperfused and hyperperfused regions on patients using a template of normal perfusion. Here we propose a new model that takes into account the first level variance and we focus on quantitative evaluation of this framework.

Section 2 presents the method developed, that focuses on two aspects : the computation of an appropriate model for normal perfusion and the selection of an adequate test to detect perfusion abnormalities at the subject level. In Section 3 the framework is evaluated based on 12 subjects diagnosed with brain tumors using a perfusion template derived from 35 healthy subjects. False positive rates (FPR) are estimated and the quality of detections are assessed.

2 Material and Methods

2.1 ASL Template: A Model of Normal Perfusion

Given $perf_i^v$, $i \in 1..N$, a set of perfusion maps in a group of N control subjects, we want to compute a model of normal perfusion $Perf^v$, referred as the perfusion template. This model is specific to each voxel v. For ease of notation, the v superscript is omitted in the following.

In [5] or [4] the perfusion signal $Perf$, is modeled by the normal distribution:

$$Perf \sim N(\mu_{pop}, \sigma_{pop}^2). \tag{1}$$

The population mean, μ_{pop}, and variance, σ_{pop}^2, are approximated by $\hat{\mu}_{pop}$ and $\hat{\sigma}_{pop}^2$, the sample mean and variance of the perfusion maps in the control group.

However, the above specified model ignores the error made when estimating the perfusion maps $perf_i$ as the average of R repeats (control minus label pairs). This error is dependent on the first level variance that has shown to carry relevant information in group fMRI studies [6]. To simplify the model, we consider that

all subjects in the template group have the same first level variance $\sigma^2_{sub,tpl}$. $Perf$ is then modeled by the normal distribution:

$$Perf \sim N(\mu_{pop}, \sigma^2_{pop} + \sigma^2_{sub,tpl}). \tag{2}$$

The first level variance, $\sigma^2_{sub,i}$, can be approximated by $\hat{\sigma}^2_{sub,i} = \frac{\hat{\sigma}^2_{rep,i}}{R}$, where $\hat{\sigma}^2_{rep,i}$ is the sample variance over the R repeats for the ith subject. The first level variance in the template group $\sigma^2_{sub,tpl}$ is estimated by $\hat{\sigma}^2_{sub,tpl} = \frac{\sum_{i=1}^{N} \hat{\sigma}^2_{sub,i}}{N}$. In this model, σ^2_{pop} can no longer be estimated as the sample variance of the perfusion maps in the control group since this estimate includes first level variance. To correct for this bias, σ^2_{pop} can be approximated by the method of moments [7] subtracting the control subjects first level variance, $\sigma^2_{sub,tpl}$, from the sample variance. Negative variance estimates are avoided by enforcing positivity of $\hat{\sigma}^2_{pop}$.

2.2   Detection of Hypoperfused and Hyperperfused Regions in a Single Subject

Given the model of normal perfusion $Perf$, we are interested in drawing conclusions about a new observation $perf_{N+1}$, for instance a patient map.

Uncorrected Probability Maps: Assuming Gaussian errors, (3) presents the best linear unbiased estimator of the patient versus group effect β, ignoring the first level variance:

$$\hat{\beta} = perf_{N+1} - \hat{\mu}_{pop}, \quad \text{Var}(\hat{\beta}) = \sigma^2_{pop} \times (\frac{1}{N} + 1). \tag{3}$$

This model is at the basis of the well-known two-sample t-test group analysis. Since $\hat{\sigma}^2_{pop}$ is estimated by the sample variance, this model is also valid in case of homoscedasticity (same first level variance for all subjects).

The best linear unbiased estimator of the patient versus group effect, β, with explicit first level variance, is expressed in (4).

$$\hat{\beta} = perf_{N+1} - \hat{\mu}_{pop}, \quad \text{Var}(\hat{\beta}) = \frac{\sigma^2_{pop} + \sigma^2_{sub,tpl}}{N} + \sigma^2_{pop} + \sigma^2_{sub,N+1} \tag{4}$$

The first level variance, $\sigma^2_{sub,N+1}$, can be approximated by $\hat{\sigma}^2_{sub,N+1} = \frac{\hat{\sigma}^2_{rep,N+1}}{R}$. Assuming that a unique patient map cannot be of better quality than the template maps, $\sigma^2_{sub,N+1}$ is constrained to be greater or equal to $\sigma^2_{sub,tpl}$.

Ignoring the error on the variance estimates, under the null hypothesis that $perf_{N+1}$ follows the template distribution, $\frac{\hat{\beta}}{\sqrt{Var(\hat{\beta})}}$ follows a standard normal distribution. To the aim of discriminating hypoperfusions from hyperperfusions, positive and negative effects are studied separately.

GLM-FDR. Most of the GLM approach has been described in the previous section. However, given the large number of voxels found in a typical ASL image a correction for multiple comparisons is required to ensure an acceptable FPR. The False Discovery Rate (FDR) algorithm insures that no more than a ratio q (usually $q = 0.05$) of detections are false positives. In GLM experiments, the data is usually pre-smoothed with a Gaussian kernel.

A Contrario: The *a contrario* approach, introduced in [8], is inspired from the Gestalt laws of perception. This theory has later been applied to medical image processing [9,10]. In our context, the *a contrario* method uses the uncorrected p-value map and takes an original approach to account for multiple comparison. The inference is done by looking into a spherical neighborhood. In the first step, the uncorrected p-value map is thresholded using a predefined set of p-values, $L = \{p_1 \dots p_T\}$, to generate a set of binary maps referred as rare event maps. For instance with a p-value of 0.001, voxels verifying $p < 0.001$ are outlined as rare events. Then, for each rare event maps, the number k^v of rare events found in a neighborhood around each voxel v is computed. The probability π_i^v of having k^v or more rare events is then estimated considering that the number of rare events in the neighborhood comes from a binomial distribution: $\pi_i^v = P(X \geq k^v)$, where $X \sim B(n, p_i)$,, $p_i \in L$ and n is the number of voxels in the studied neighborhood. For the sake of multiple comparisons correction, a number of false alarm (NFA) map is then computed by $\text{NFA}_v = V\,T\,\min(\pi_i^v)$ where V is the total number of voxels. $V\,T$ therefore corresponds to the total number of tests. Voxels verifying $NFA < 1$ are outlined as detection. Due to the neighborhood constraint, a voxel can be detected both as hyperperfused and hypoperfused. To avoid this confusing situation, a voxel cannot be outlined as hyperperfused (rep. hypoperfused) if its value is smaller (resp. greater) than the template mean. In this paper we worked with the following set of p-values $P = \{0.001, 0.005\}$.

3   Results

3.1   Data

Data: 36 healthy volunteers and 14 patients diagnosed with brain tumors were involved in this study. Data acquisition was performed on a 3T Siemens Verio MR scanner with a 32-channel head-coil. The imaging protocol included a 3D T1-weighted anatomical sequence (T1) (TR: 1900 ms, TE: 2.27 ms, resolution: $1 \times 1 \times 1$ mm^3), a PICORE Q2TIPS sequence with crusher gradients (TR: 3000 ms, TE: 18 ms, resolution: 3×3 mm^2, slice thickness: 7 mm, TI: 1700 ms, TI$_{wd}$: 700 ms, R = 60). In addition to these sequences, the patients also underwent a 3D T1 post gadolinium (T1-Gd) sequence (TR: 1900 ms, TE: 2.27 ms, resolution: $1 \times 1 \times 1$ mm^3) and a T2 FLAIR sequence (TR: 9000 ms, TE: 90 ms, resolution: 0.69×0.69 mm^2, slice thickness: 4 mm). Three subjects (1 control, 2 patients) were excluded for abnormally low signal.

Pre-processing: Image pre-processing was performed using the Matlab tool-box SPM8 (Wellcome Department of Imaging Neuroscience, University College, London). The anatomical image of each subject was segmented using the unified segmentation. For each subject, an anatomical brain mask was created, excluding voxels with less than 50% of brain tissue in subsequent statistical analyses. A six-parameter rigid-body registration of the ASL volumes was carried out in order to reduce undesired effects due to subject motion. Coregistration on grey matter map was then performed based on mutual information. The average of unlabeled volumes was used to estimate the geometrical transformation to apply to each volume. The 60 unlabeled and labeled ASL volumes were pair-wise subtracted and averaged in order to obtain a perfusion weighted map per subject. A standard kinetic model [11] was then applied in order to get ASL CBF.

In order to account for inter-subject variations in CBF, each map was normalised by the mean perfusion value computed from all voxels containing more than 70% of GM [5]. In patients, tumorous tissue was excluded from the calculation. Spatial normalisation parameters estimated during the segmentation step were then applied to the T1 and ASL CBF map in order to normalise the subjects into the ICBM template space [12].

3.2 Influence of First Level Variance in the Model of Normal Perfusion

Quantitative Comparison: We compared the basic model of normal perfusion, presented in (1), to the model with explicit first level variance estimation, outlined in (2), by means of FPR estimation. FPR estimates were computed by leave-one-out cross-validation on the control group using the tests specified in (3) and (4). Given the absence of correction for multiple comparisons the theoretical FPR was therefore equal to the selected p-value: p=0.05.

With the basic model (1) the FPR was 6.5%, outlining the error made by ignoring first level variance. With the model including first level variance (2), the FPR was reduced to 4.3%. Though the difference was not strong, the FPR of the model including first level variance was closer to the theoretical FPR. Moreover, benefits could potentially be larger in the study of patients presenting more pronounced patterns of noise. As an example, the following subsection illustrates the benefits in a single subject presenting strong ghosting artefacts.

Qualitative Comparison: Figure 1 illustrates the benefits of the explicit modeling of first level variance. The ghosting and motion artefacts, outlined by white arrows, indeed corresponds to regions of high standard deviation. Including the first level variance in the model reduces the artefactual detections and enables the comparison of individuals with different patterns of noise.

3.3 Validation of Detections

Metrics Definition: Quantitative evaluation of the detections and comparison between detection methods are challenging tasks. This is mainly because, like in
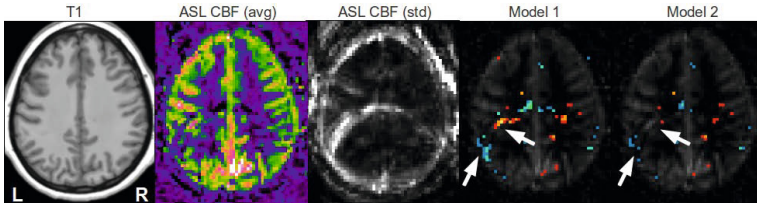
Fig. 1. False positive detections in a control subject perfusion map presenting ghosting artefacts. From left to right: T1, ASL CBF, standard deviation of ASL CBF, ASL CBF with hypo- (winter color-map) and hyper-perfusions (hot color-map) overlaid for the basic model (Model 1) and for the model including first level variance (Model 2)

many other medical imaging problems, the ground truth is not clearly stated. We chose to evaluate this framework on patients diagnosed with tumor pathology because perfusion abnormalities are better understood in this context. Based on clinical knowledge, we set the following rules to get an idea of the specificity and sensitivity of the methods. It is however important to keep in mind that in the absence of a known delineation of the actual hyperperfusion and hypoperfusion, true specificity and sensitivity cannot be calculated:

– **pseudo-specificity:** According to clinical knowledge, in the absence of metastasis, the perfusion abnormalities should be confined to the affected tissue (tumor and oedema) identifiable on T1-Gd and T2. The proportion of the non-affected tissue undetected was used as a measure of pseudo-specificity.
– **pseudo-sensitivity:** Though hyperperfused regions can be extended out of the T1-Gd hyper-signal, these hypersignals are nevertheless indicative of the presence of hyperperfusions [3]. The proportion of the hyper-signal in T1-Gd detected as hyperperfusion was used as a measure of pseudo-sensitivity.

Tumor segmentation was done using a semi-automated method based of the T2 and T1-Gd images and visually inspected by an expert neuro-radiologist. One subject was excluded in the pseudo-specificity calculation because its T1-Gd hyper-signal was located in the cerebellum out of the coverage of the template.

Quantitative Evaluation: Table 1 presents the pseudo-specificity and pseudo-sensitivity obtained with different parameters. Based on these criteria we cannot identify a method as being definitely better than the other. This study however outlines the different behaviors of the two methods. The GLM with FDR correction appears to be more conservative with a smaller FPR at the cost of false negatives. According to these measures, the choice of the method will therefore depend on the application as a trade-off between specificity and sensitivity.

Qualitative Comparison: Though the quantitative analysis did not outline significant difference between the *a contrario* and the GLM methods, there are however strong differences when looking at the detections patient by patient. We chose three subjects to illustrate this argument. Figure 2 presents the detections obtained by GLM with smoothing kernel of 8 and *a contrario* with sphere radius

Table 1. Pseudo-sensitivity and pseudo-specificity of detections for GLM with different full width at half maximum w (in mm^3) smoothing kernels and *a contrario* methods with different sphere radii r (in voxels)

	GLM					*a contrario*	
	$w = 2$	$w = 4$	$w = 6$	$w = 8$	$w = 10$	$r = 1$	$r = 2$
$pseudo-sensitivity$	0.29	0.31	0.32	**0.33**	0.34	0.37	**0.53**
$pseudo-specificity$	0.98	0.97	0.96	**0.95**	0.94	0.96	**0.89**

of 2 which are the parameters leading to the best visual results for each method. Subject 1 is affected by a tumor in the right frontal lobe. The hypoperfusion of the oedema is detected by both the GLM and the *a contrario* methods. However only the latter is able to outline the central hyperperfusion. Subject 2, presents a left temporal glioblastoma characterised by a ring shaped hypersignal. While the *a contrario* approach detects most of the hypersignal ring, the GLM detection is restricted to the area presenting the highest level of hyperperfusion. Subject 3 suffers from a meningioma of both occipital lobes, the hyperperfusion is correctly detected by both methods. The *a contrario* method detects additional hypoperfusions in the oedema. Both methods detect an artefactual hyperperfusion, due to motion, in the frontal lobe. The *a contrario* approach also detects artefactual hypoperfusions due to motion in the occipital lobe.
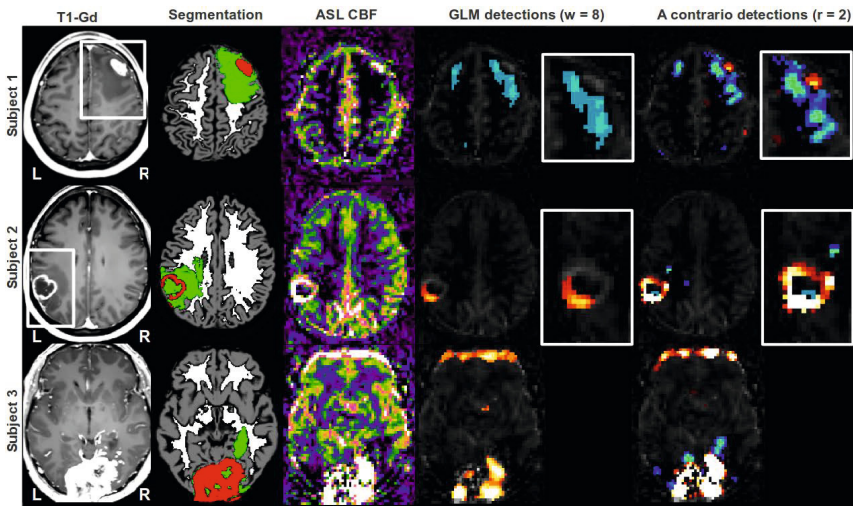
Fig. 2. Detection of perfusion abnormalities based on GLM and *a contrario* methods on 3 representative patients with brain tumors. From left to right: T1-Gd, Segmentation in GM (grey), WM (white), oedema and necrosis (green) and T1-Gd hypersignal (red), ASL CBF with hypoperfusions (winter color-map) and hyperperfusions (hot color-map) overlaid, close-ups in the regions outlined by a white frame in the T1-Gd.

4   Conclusion

We have presented a comprehensive framework for the detection of brain perfusion abnormalities in individual patients by comparison to a template of healthy subjects. We outlined the interest of modeling the first level variance in the perfusion models. In future work, this might ease the analysis of patient data characterised by a higher level of motion and related artefacts. It also opens the field to comparison between ASL sequences displaying different patterns of noise. We applied this model to 12 patients suffering from brain tumors and compared an original method, the *a contrario* approach, to the classical GLM with FDR correction. Quantitative analysis outlined the conservativeness of the GLM, leading to a low FPR at the cost of more false negatives. In radiological practice, the FPR is however less pregnant since false positives can easily be ruled out by a clinician. In a qualitative analysis we pointed out the benefits of the *a contrario* approach: a better conservation of the hypoperfusions and hyperperfusions boundaries and a greater sensitivity. This increase in sensitivity might be crucial in the study of pathologies presenting more subtle patterns of abnormal perfusion such as multiple sclerosis.

References

1. Detre, J., Leigh, J., et al.: Perfusion imaging. MRM 23, 37–45 (1992)
2. Alsop, D.C., Detre, J.A., Grossman, M.: Assessment of Cerebral Blood Flow in Alzheimers Disease by Spin-Labeled Magnetic Resonance Imaging. Annals of Neurology 47(1), 93–100 (2000)
3. Weber, M.A., Zoubaa, S., et al.: Diagnostic performance of spectroscopic and perfusion MRI for distinction of brain tumors. Neurology 66(12), 1899–1906 (2006)
4. Maumet, C., Maurel, P., et al.: A contrario detection of focal brain perfusion abnormalities based on an ASL template. In: ISBI, pp. 1176–1179 (2012)
5. Petr, J., Ferre, J.C., et al.: Construction and evaluation of a quantitative arterial spin labeling brain perfusion template at 3t. In: ISBI, pp. 1035–1038. IEEE (2011)
6. Beckmann, C.F., Jenkinson, M., Smith, S.M.: General multilevel linear modeling for group analysis in FMRI. NeuroImage 20(2), 1052–1063 (2003)
7. Chen, G., Saad, Z.S., et al.: Fmri group analysis combining effect estimates and their variances. NeuroImage 60(1), 747–765 (2012)
8. Desolneux, A., Moisan, L., Morel, J.: A grouping principle and four applications. TPAMI 25(4), 508–513 (2003)
9. Aguerrebere, C., Sprechmann, P., et al.: A-contrario localization of epileptogenic zones in SPECT images. In: ISBI, pp. 570–573. IEEE (2009)
10. Rousseau, F., Blanc, F., et al.: An a contrario approach for outliers segmentation: Application to Multiple Sclerosis in MRI. In: ISBI, pp. 9–12 (2008)
11. Buxton, R.B., Frank, L.R., et al.: A general kinetic model for quantitative perfusion imaging with arterial spin labeling. MRM 40(3), 383–396 (1998)
12. Crinion, J., Ashburner, J., et al.: Spatial normalization of lesioned brains: performance evaluation and impact on fmri analyses. NeuroImage 37(3), 866–875 (2007)

Automated Colorectal Cancer Diagnosis
for Whole-Slice Histopathology

Habil Kalkan[1,3,⋆], Marius Nap[2], Robert P.W. Duin[1], and Marco Loog[1]

[1] Pattern Recognition Laboratory, Delft University of Technology, The Netherlands
[2] Atrium Medical Center, Heerlen, The Netherlands
[3] Department of Computer Engineering, Suleyman Demirel University, Turkey

Abstract. In this study, we propose a computational diagnosis system for detecting the colorectal cancer from histopathological slices.The computational analysis was usually performed on patch level where only a small part of the slice is covered. However, slice-based classification is more realistic for histopathological diagnosis. The developed method combines both textural and structural features from patch images and proposes a two level classification scheme. In the first level, the patches in slices are classified into possible classes (adenomatous, inflamed, cancer and normal) and the distribution of the patches into these classes is considered as the information representing the slices. Then the slices are classified using a logistic linear classifier. In patch level, we obtain the correct classification accuracies of 94.36% and 96.34% for the cancer and normal classes, respectively. However, in slice level, the accuracies of the 79.17% and 92.68% are achieved for cancer and normal classes, respectively.

1   Introduction

Colorectal cancer is the third most common cancer in both men and women world-wide and is the third leading cause of cancer-related deaths in the Western world [1]. For 2012, 103,000 colon cancer cases and 51,000 colon cancer related deaths are predicted for the United States. Like for many other types of cancer, histopathological analysis is accepted as the gold standard for malignancy diagnosis [2]. The analysis of these data, however, is performed visually and the detection and grading of the suspect tissue may show variability depending on the experience and awareness of the experts. Therefore, various studies have been performed into the development of computer-aided diagnosis systems (CAD) to improve the ability of pathologists at discriminating between malignant and benign tissue.

As the presence and grade of malignancy in the cell tissue is strongly correlated to histological structures like, lumina, stroma, nuclei and glands, most CAD research is focused on structural shape changes. A gland, which has lumina at the center and is surrounded by the stroma and nuclei, loses its regular structure

---

with progressing malignancy [3,4]. In addition to glands, the number of nuclei and also the arrangement of the nuclei in the tissue can have diagnostic significance for some kind of malignancy in histopathology [5,6]. Numerous methods have been used to detect the nuclei in histopathology images and various automated detection results were compared against manual segmentation [7,8]. It has, however, been indicated that the histological objects in the tissue may not need to be perfectly detected for good tissue classification when a comprehensive set of features is available [9]. Boucheren et. al [10] achieved good classification results on breast cancer histopathology with the imperfectly segmented nuclei. In addition to glands and nuclei, the malignancy drastically changes the density of lumina and stroma in tissue. The relevance of texture information on the cytoplasm and stroma structures was reported for classification. Sertel et.al [11] evaluated Haralick and Local Binary Pattern (LBP) features categorizing the nerve histopathology images into *stroma rich* and *stroma poor* classes for nervous cancer prognosis.

CAD in histopathological image analysis [11] is, however, still at the evolving level compared to CAD in radiology [12]. Besides, recent studies in histopathology image analysis were mostly focused on prostate and breast cancer [6,5] detection but very limited study was encountered on computational diagnosis of colorectal cancer [13,14]. In general, the computational histopathological image analysis is performed on patch level rather than slice-level due to their relatively large size [6,13,3,8]. Serter et.al [11] proposed a multi-scale analysis system to classify the nervous tissue images into stroma-rich and stroma-poor regions for neuroblastoma cancer detection. In colon histopathology, Ficsor et. al [14] extracted the glandular and nuclei structure in patches, then combined the patch-level statistics to get slice-level features for classification of the Crohn's diseases, ulcerative colitis and aspecific colitis. However, slice-based approach proposed in [14] is not appropriate for cancer diagnosis. Because nonmalignant regions in a tissue may suppress the malignant regions in a slice if global statistics is used. To the best of our knowledge we are the first to present both a scheme that performs a diagnosis at the level of the full image and a method that focuses on colorectal cancer.

We propose a two level classification scheme for classifying full colon cancer histopathological images, which we refer to as slices. These very large sized slices are built up of sub-images, which we refer to as patches. In the first level, we use patch-labeled tissue images to generate a four class classifier with classes *normal*, *cancer*, *adenomatous* and *inflamed* (see Fig.2). Subsequently, patches in slices are classified into these four classes. Each slice is then represented by a feature vector from the patch classification results based on which the entire slice can be labeled and our procedure can be evaluated. Further details on the data employed together with the used feature extraction, selection, and classification algorithms are presented in Section 2. Experimental results and conclusions are given in Sections 3 and 4, respectively.

2   Materials and Methods

2.1   Data

A total of 120 H&E stained colon biopsy slices from 96 different patients were collected and scanned by high resolution camera at Atrium Medical Center, Heerlen. The original slices (sized about 70,000x120,000) consist of equal sized patches of 1024x1024 pixels. Each slice may include different numbers of patches that display actual tissue (between 200 and 6000) depending on the size of the original biopsy etc. Slices were labeled at two different levels: either at *slice-level* or at patch-level. 55 of the slices are used for *patch-based* labeling where each individual patch was assigned to one of the four primary (*normal, cancer, adenomatous, inflamed*) and two secondary classes (*unknown, inappropriate*) by a pathologist (see Fig.2). The unknown class is for the patches which the pathologist is not sure and inappropriate is for the patches which are not appropriate for analysis due to the imaging problems such as camera focus. We consider the primary classes only, which include 6134, 2503, 2261 and 2967 patches, respectively.
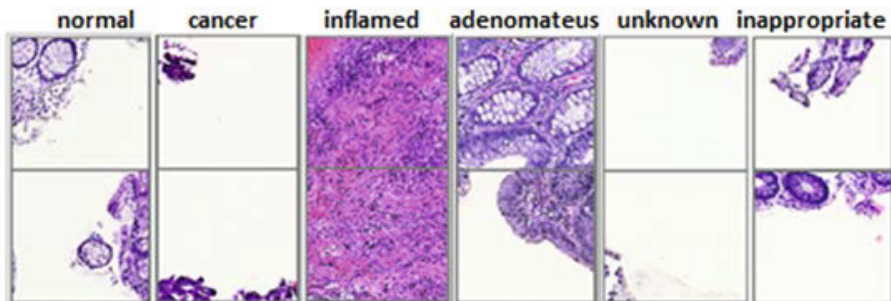
Fig. 1. Patch images from *patch-based* dataset for six different classes

The remaining 65 slices were labeled in a *slice-based* manner where 24 of them were assigned to the *cancer* and 41 of them were assigned to the *normal* class. The patches in these 65 slices were not labeled individually. The *slice-based* dataset is more challenging for pattern recognition tasks but more realistic for analysis in pathology. A pathologist assigns the slices to one class although it may include patches from different classes. For example, a *cancer* labeled slice may include patches from *non-cancer* classes as well as it includes *cancer* patches, though the opposite situation is unlikely. To limit the computation times, we randomly selected a number of patches from every slices as a representation, with a maximum number of 300. Taking more than 300 patches hardly leads to improved performance as the necessary statistics are already estimated well.

2.2   Structural Features

In colon tissue, glands may have irregular shapes even in nonmalignant slices (Figure 1). Therefore, we ignored the shape of the glands and focused on the

density of the structures and nuclei in tissue. We divide the tissue image into background, stroma, lumina, and nuclei (Fig. 2b) using k-means clustering. We initialize the center of the clusters with the empirically detected RGB triplet values $\{(0,0,0),(0.29,0.19,0.4),(0.91,0.91,0.94),(0.7,0.5,0.72)\}$ and iteratively refine the clusters until no more assignment is observed. The background segment (Fig. 2b) is used for masking. Next, the RGB image is transformed to gray levels, which are normalized. These normalized images are then used for nuclei detection.
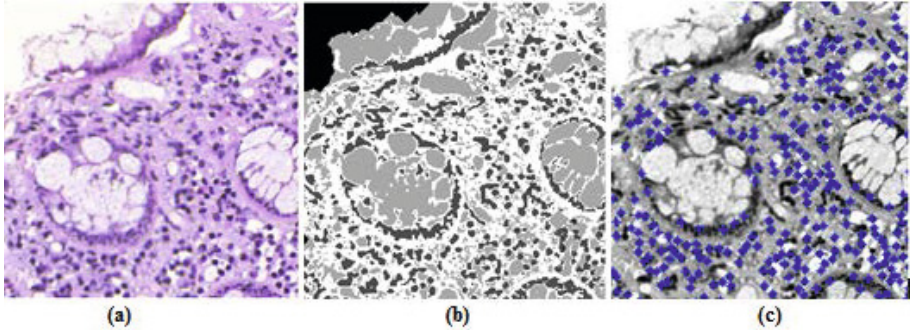
Fig. 2. a) a tissue patch from normal class, b) clustered into four segments, c) detected nuclei (shown in blue star)

Nuclei are detected by Laplacian filtering. Because of slight variations in scale, we used Lindeberg's blob detection algorithm with automatic scale selection over a small scale range to find the nuclei [15]. Using this algorithm, we can detect a larger number of nuclei around and between the glands (Fig.2c). With this procedure, the connected nuclei around the glands, which are concatenated to each other, are missed however. The nuclei in patches could be better segmented using more advanced techniques. The findings in [9,10] indicate, however, that imperfect detections like the ones we employ do not necessarily hamper good classification results.

The number of nuclei in the tissue and their arrangement has significant importance for tissue classification. To capture this, we divide every patches into 16 sub-patches, considering only sub-patches that contain tissue. We then extract local shape features for each sub-patch, such as the number of nuclei per tissue, the ratio of each individual segment (stroma, nuclei and lumen) to the whole tissue, and pairwise ratios of individual segments. Then the mean and variance of these local features are evaluated to yield 19 structural features to represent a patch.

2.3 Texture Features

H&E staining in pathology colors the structures in tissue with different colors (i.e. blue-purple, pink etc). However, we observed a number of improperly stained

tissues which the colors are inappropriately distributed in RGB space. Therefore, we moved to HSV space for texture analysis and extracted texture features from H, S, V components together with normalized gray value components.

Haralick Features. We extract four of the second order statistical texture [16] features. We first evaluate gray level-concurrence matrix $M(\theta) \in \mathbb{R}^{NxN}$ in four different directions where $\theta \in \left\{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}\right\}$. Using this matrices, we extract the four Haralick features; homogeneity, contrast, energy and correlation for each directions to yield 16 features.

Gabor Filter Features. We construct a set of Gabor filters $G(f, \theta)$ [17,5] for orientation parameters $G(f, \theta)$ and frequency parameters $f \in \{k, 2k, ..., 8k\}$ where k is selected to be 1/16 to produce filter within one period. With the inclusion of the filter for the parameter (0,0) we obtain 65 different Gabor filters. We filter the tissue image with the obtained filters and the average, variance and minimum-to-maximum ratio of these filtered images are used as features to yield a total of 195 Gabor features for the given image.

Color Channel Histograms. The existence of malignancy and also the level of malignancy alter the structures in tissue and this significantly the effect the color distribution after H&E staining. Therefore, color channel histograms of R, G and B component of the raw tissue image are obtained after removing the background pixels. We evaluate 32 bin histograms for each channel to yield a total of 96 features for each raw image.

2.4   Feature Selection

A total of 1287 (1269 texture + 19 structural) features are extracted using the algorithms defined in previous section. Not all features, however, may be relevant for the classification of patches and having too many of them typically has a negative effect on the classification accuracy. Besides, it is computationally demanding to extract large number of features for large data set like high resolution microscopy images. Therefore, we chose to reduce the data dimensionality by simply ranking the features according to their discrimination potential using the sum of Euclidean distances between all four class means [18]. To decide on the number of features to retain, features are incrementally included and tested for classification performance. The optimal number of features is simply that number that gives the lowest estimated classification error.

2.5   Classification and Slice Level Fusion

Using the reduced feature set, a k-NN classifier is trained for four-class patch classification in which the neighborhood parameter k is optimized by means of leave-one-out cross validation. As our procedure should operate at a slice level, rather than patch level, we propose a second level fusion step for slices based on the patch based classification scores obtained on every slice. Using the k-NN

classifier, this second step collects all the posterior probabilities from the (up to) 300 patches that represent every slice and calculates the average posterior for all of the four primary classes. These averaged posteriors can now be taken as new feature vectors that represents a full slice. Using these features, a logistic regression is used to come to a final diagnosis for a whole slice. We note that the initial patch classifier has been trained on the first part of the data only. None of the 65 test images were involved in this. The second stage classifier is also evaluated by means of leave-one-out.

3   Experimental Results

3.1   Patch-Based Classification

The *patch-labeled* dataset with *cancer* and *normal* classes is randomly divided into test and train sets and the best 155 number of features are selected considering their individual Euclidean distances between the classes. A logistic-linear classifier is used for classification and the mean classification error is considered for evaluation. The patches in colon histopathology may have different tissue ratio (tissue area/ patch area). In order to analyze the effect of tissue ratio for classification we restricted the lower bound of the tissue ratio for specific values and perform classification only with the patches whose tissue rate is higher than the defined values (Fig. 3).
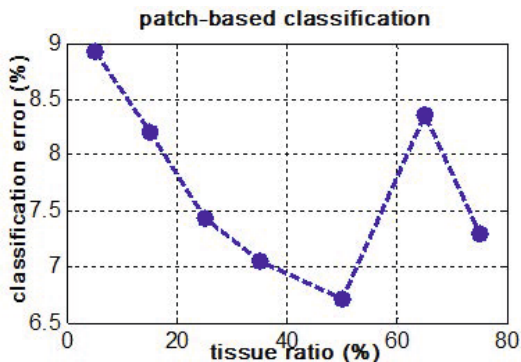
Fig. 3. Minimum classification error rates with respect to the lower bound in tissue ratio

It is observed that the error rate decreases if we move the lower bound to 50% but then starts to increase. Increasing the tissue ratio inversely decreases the number of objects in dataset and the decrease of objects in dataset explains the increase on the error curve. However, We fix the tissue ratio to 25% for further analysis and obtain the correct classification rates of 94.36% and 96.34% for the cancer and normal classes, respectively.

3.2   Slice-Based Classification

For slice-based classification, the entire objects in patch-based dataset are used for training a k-nn classifier with four main classes (*normal*, *cancer*, *adenomatous* and *inflamed*). Then, the unlabeled patches in each slice are classified into these four classes and feature vector representing the entire slice are obtained based on the patch-based classification. The slices are then classified by using leave-one out principle using a logistic linear classifier and correct classification accuracies of 79.17% and 92.68% are achieved for the cancer and normal slices, respectively.

The results in slice-based classification give rise to an 87.69% correct classification accuracy which is lower than the accuracy obtained in *patch-based* classification. The accuracy of testing is measured by the area under ROC curve (AUC) and we obtain a AUC value 0.90 for the slice-based classification.

4   Conclusion

An automatic whole-slice based histopathological image analysis method is proposed to determine the colon cancer. In the proposed approach, a two level classification scheme is proposed to handle the patch-based and slice-based recognition based on the texture and structural features. In the first level, path-based classification is performed between *cancer* and *normal* classes and the classification accuracies of 94.36% and 96.34% are achieved for each class, respectively. The tissue ratio significantly affects the classification results. Eliminating the patches with, respectively, less tissue ratio decreases the classification error. However, marginal elimination should be avoided because malignancy may exist at even small tissues. The obtained patch-based classification results are promising and it is more or less than the results obtained in literature. However, the main contribution of the study is on the whole-slice classification which is more realistic for histopathology. Because, malignancy does not spread homogeneously in the tissue and the patches acquired with high resolution scanner may be in different diagnostic labels. The whole-slice based classification task is achieved in two levels; first the patches in slices are patch-based classified into *normal*, *cancer*, *adenomatous* and *inflamed* classes and then the distribution of the patches to these classes are considered to get a higher level feature vector representing the slice. Then the slices are classified into two *cancer* and *normal* classes where 19 of the 24 *cancer* and 38 of the 41 *normal* slices are correctly classified which makes a 87.69% mean classification accuracy in slice level.

References

1. Cancer Facts and Figures 2012. American Cancer Society, Atlanta (2012)
2. Rubin, R., Strayer, D., Rubin, E., McDonald, J.: Rubin's Pathology: Clinicopathologic Foundations of Medicine. Lippincott W. & W., Baltimore (2007)
3. Farjam, R., Soltanian-Zadeh, H., Jafari-Khouzani, K., Zoroofi, R.A.: An Image Analysis Approach for Automatic Malignancy Determination of Prostate Pathological Images. Cytometry Part B (Clinical Cytometry) 72B, 227–240 (2007)

4. Gleason, D.F.: Histologic grading of prostate cancer: A perspective. Hum. Pathol. 23, 273–279 (1992)
5. Doyle, S., Agner, S., Madabhushi, A., Feldman, M., Tomaszewski, J.: Automated grading of breast cancer histopathology using spectral clustering with textural and architectural image features. In: IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pp. 496–499 (2008)
6. Axelrod, D.E., Miller, N.A., Lickley, H.L., Qian, J., Christens-Barry, W.A., Yuan, Y., Fu, Y., Chapman, J.A.: Effect of quantitative nuclear image features on recurrence of ductal carcinoma in situ (dcis) of the breast. Cancer Informatics 4, 99–109 (2008)
7. Ibba, A., Duin, R.P.W., Loog, M.: Supervised localization of cell nuclei on TMA images. In: IEEE Computer Society Workshop on MMBIA (2012)
8. Naik, S., Doyle, S., Agner, S., Madabhushi, A., Tomaszewski, J., Feldman, M.: Automated gland and nuclei segmentation for grading of prostate and breast cancer histopathology. In: ISBI Special Workshop on Comput. Histopathology, pp. 284–287 (2008)
9. Boucheron, L.E.: Object- and Spatial-Level Quantitative Analysis of Multispectral Histo-pathology Images for Detection and Characterization of Cancer, Ph.D. dissertation, Univ. of California Santa Barbara, Santa Barbara, CA (2008)
10. Boucheron, L.E., Manjunath, B.S., Harvey, N.R.: Use of imperfectly segmented nuclei in the classification of histopathology images of breast cancer. In: IEEE International Conference on Acoustics Speech and Signal Processing, pp. 666–669 (2010)
11. Sertel, O., Kong, J., Shimada, H., Catalyurek, U., Saltz, J., Gurcan, M.N.: Computer-aided prognosis of neuroblastoma on whole-slide images: Classification of stromal development. Pattern Recognit. 42, 1093–1103 (2009)
12. Gurcan, M.N., Boucheron, L.E., Can, A., Madabbushi, A., Rajpoot, N.M., Yener, B.: Histopathological image analysis: A review. IEEE Rev in Bio. Eng. 2 (2009)
13. Esgiar, A.N., Naguib, R.N.G., Sharif, B.S., Bennett, M.K., Murray, A.: Fractal Analysis in the Detection of Colonic Cancer Images. IEEE Trans. on Infor. Tech. in Biomedicine 6(1) (2002)
14. Ficsor, L., Varga, V.S., Tagscherer, A., Tulassay, Z., Molnar, B.: Automated Classification of Inflammation in Colon Histological Sections Based on Digital Microscopy and Advanced Image Analysis. Cytometry Part A 73A, 230–237 (2008)
15. Lindeberg, T.: Detecting salient blob-like image structures and their scales with a scale-space primal sketch: A method for focus-of-attention. International Journal of Computer Vision 11(3), 283–318 (1993)
16. Haralick, R.M.: Statistical and structural approaches to texture. Proc. IEEE 67(5), 786–804 (1979)
17. Ilonen, J., Kamarainen, J.-K., Paalanen, P., Hamouz, M., Kittler, J., Kälviäinen, H.: Image feature localization by multiple hypothesis testing of Gabor features. IEEE Transactions on Image Processing 17(3), 311–325 (2008)
18. Theodoridis, S., Koutroumbas, K.: Pattern Recognition. Academic Press, London (2008)

Patient-Adaptive Lesion Metabolism Analysis by Dynamic PET Images

Fei Gao[1], Huafeng Liu[2,1], and Pengcheng Shi[1]

[1] Golisano College of Computing and Information Sciences, Rochester Institute of Technology, Rochester, NY, 14623, USA
[2] State Key Laboratory of Modern Optical Instrumentation, Zhejiang University, Hangzhou, 310027, China

Abstract. Dynamic PET imaging provides important spatial-temporal information for metabolism analysis of organs and tissues, and generates a great reference for clinical diagnosis and pharmacokinetic analysis. Due to poor statistical properties of the measurement data in low count dynamic PET acquisition and disturbances from surrounding tissues, identifying small lesions inside the human body is still a challenging issue. The uncertainties in estimating the arterial input function will also limit the accuracy and reliability of the metabolism analysis of lesions. Furthermore, the sizes of the patients and the motions during PET acquisition will yield mismatch against general purpose reconstruction system matrix, this will also affect the quantitative accuracy of metabolism analyses of lesions. In this paper, we present a dynamic PET metabolism analysis framework by defining a patient adaptive system matrix to improve the lesion metabolism analysis. Both patient size information and potential small lesions are incorporated by simulations of phantoms of different sizes and individual point source responses. The new framework improves the quantitative accuracy of lesion metabolism analysis, and makes the lesion identification more precisely. The requirement of accurate input functions is also reduced. Experiments are conducted on Monte Carlo simulated data set for quantitative analysis and validation, and on real patient scans for assessment of clinical potential.

1 Introduction

Dynamic Positron Emission Tomography (dPET) is a molecular imaging technique that is used to monitor the spatiotemporal distribution of a radiotracer *in vivo* and enables cellular level metabolism analysis in clinical routine. dPET provides a good promise for quantitative lesion metabolism analysis to help identify lesions. However, due to poor statistical properties of the measurement data in low count dynamic PET acquisition and disturbances from surrounding tissues, identifying small lesions inside the human body is still a challenging issue. Furthermore, the mismatch between general purpose models and patient size/motions makes the situation even worse.

Quantitative kinetic analysis of radiotracer uptakes requires the reconstruction of kinetic parameters[1–3]. The mainstream is statistical reconstruction algorithms, however, whose quality is determined by the accuracy of sophisticated

system probability matrix (SM). Many efforts have been devoted to improve the accuracy of SM [4–7]. However, the ideal SM is almost impossible to obtain under practical conditions. The general purpose SM also could not compensate different sizes of patients and the motions during acquisition, which will decrease the accuracy of reconstructions. Furthermore, the reconstruction of dynamic PET image sequences, whose poor temporal resolution, insufficient photon counts, more complicated data corrections and poor statistical properties of measurement data also requires a more accurate SM.

In this paper, we present a dynamic PET metabolism analysis framework by defining a patient adaptive system matrix to improve the lesion metabolism analysis. Both patient size information and potential small lesion information are incorporated by simulations of phantoms of different sizes and individual point source responses[8–10]. Experiments of 90 studies are conducted using 15 phantoms of different sizes based on Zubal thorax phantom. Each experiment has randomly generated motions and a lesions in lung. Both true lesion and false lesion cases are studies. We also analyze the results using input functions of different accuracies. Our method shows obvious improvements in identifying lesions (including sizes, true/false situations, metabolism rates), and reduces the requirement of the accuracy of input functions. An experiment based on real patient scans is also conducted for assessment of clinical potential.

2   Method

2.1   Tracer Kinetics

Dynamic PET imaging provides the opportunities to perform lesion metabolism analysis by using compartmental models to quantitatively describe regional radiotracer kinetics. A typical three compartmental model (Phelps 4K model) can be mathematically represented by a set of ordinary differential equations [11].

$$\frac{dC_F(t)}{dt} = k_1(t)C_P(t) + k_4(t)C_B(t) - (k_2(t) + k_3(t))C_F(t) \tag{1}$$

$$\frac{dC_B(t)}{dt} = k_3(t)C_F(t) - k_4(t)C_B(t) \tag{2}$$

where C_P(pmol/ml) is arterial concentration of injected radiotracer, C_F is the free and non–specific binding ligands, C_B is the specific binding tracers in tissues. Parameters k_1, k_2, k_3 and $k_4(min^{-1})$ are first-order rate constants specifying radiotracer transport rates. The general PET measurement equation is $Y = DX + e$. With the compartment model introduced, the activity distribution X should be the combination of C_F, C_B, C_P and fractional volume of blood f_v

$$X(t) = (1 - f_v)(C_F(t) + C_B(t)) + f_v C_P(t) \tag{3}$$

C_F, C_B are the functions of kinetic parameters, if defining $\kappa = \{k_1, k_2, k_3, k_4\}$

$$Y(t) = DX(\kappa, t) + e(t) \tag{4}$$

2.2   System Matrix Derived from Supervised Learning

Statistical reconstruction requires a well modeled SM, which directly determines the accuracy of reconstruction results. The SM D is extended to include 2 parts, D_1 is a SM generated from geometry information and physical phenomena, and will account for sizes and motions of different patients, D_2 is an additional SM generated from point source responses. D_1 and D_2 are full size SM, and combined together by weighting matrices w_1 and w_2 according to the anatomical information of patients. This effort makes the SM more patient adaptive. The measurement equation is extended from Eqn.4 to be

$$Y(t) = \begin{bmatrix} w_1 \ w_2 \end{bmatrix} \begin{bmatrix} D_1 \\ D_2 \end{bmatrix} X(\kappa, t) + e(t) \tag{5}$$

w_1, w_2, D_1, D_2 are updated by supervised learning. Training sets are provided by Monte Carlo simulations using GATE toolbox. Correspondingly, 2 series of simulations are performed, one is performed with human thorax phantom of different sizes, and the other is done by point source response inside a thorax phantom of normal size. Denoting the activity concentrations as $X = \{x_1, x_2, \cdots x_n\}$ and measurement datasets as $Y = \{y_1, y_2, \cdots x_n\}$. n is the number of training sets, and every dataset is a dynamic data sequence related to time t. For simplification of expression, Eqn.5 is written as $Y(t) = D'X(k, t) + e(t)$. Since ADALINE has been proved to be simple yet successful for updating SM in[7], we also adopt ADALINE for our SM training here. The initialization of D_1 and D_2 are the SMs generated with uniform cylindrical phantom. The update procedure by ADALINE using back-propagation and least mean square error is:

$$\hat{y}_m(t) = D'_m X(k, t) + e_m(t) \qquad \delta_k(t) = Y(t) - \hat{y}_m(t) \tag{6}$$
$$D'_{m+1}(t) = D'_m(t) + 2L\delta_m X^T(t) \qquad e_{m+1}(t) = e_m(t) + 2L\delta_m(t) \tag{7}$$

where m is the iteration step of training, and L is the learning rate. After defining a precision level of learning ϵ,

$$D' \qquad \textbf{subject to} \qquad \begin{cases} Y(t) - \hat{y}_m(t) < \epsilon \\ \hat{y}_m(t) - Y(t) < \epsilon \end{cases} \tag{8}$$

the weighting matrices w_1 and w_2 will be obtained when convergence is achieved.

2.3   Parameter Reconstruction of Dynamic PET

The kinetic model and image reconstruction are combined in one equation, the log likelihood function can be derived with measurement data $y(t)$ as

$$\mathcal{L}(y|\kappa) = \sum_t y(t) \log \bar{y}(\kappa, t) - \bar{y}(\kappa, t) \tag{9}$$

$$\textbf{where} \quad \bar{y}(\kappa, t) = D'x(\kappa, t) + e(t) \tag{10}$$
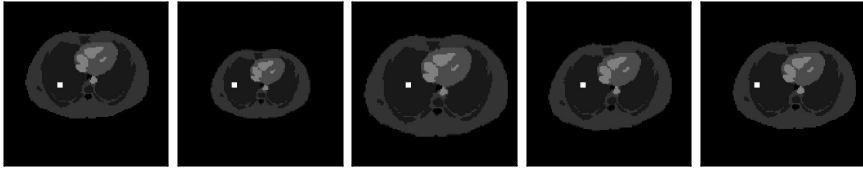
Fig. 1. First 5 phantoms

(a)   (b)   (c)
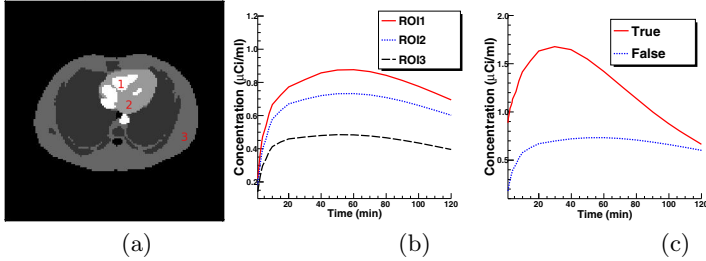
Fig. 2. (a) Standard Zubal Phantom; (b) TAC curves of 3 ROIs indicated in (a); (c) TAC curves of true lesion and false lesion

$$\hat{\kappa} = \arg\max \Phi(\kappa), \qquad \Phi(\kappa) = \mathcal{L}(y|k) - \beta \mathcal{U}(k). \tag{11}$$

where $\mathcal{U}$ is the penalty regularization term with parameter β controlling resolution/noise tradeoff. Eqn.11 is solved by a paraboloidal surrogates algorithm in [12]. Since the parameter reconstruction has a higher data dimensionality/ freedom, we also define the evaluation of a student's t-distribution hypothesis test to determine their statistical differences among iterations. By selecting Region of Interest (ROI), calculate $t = \frac{|\bar{x}_m - \bar{x}_{m+1}|}{\sigma}$, where $\sigma = (\frac{var_m + var_{m+1} - 2cov_{m,m+1}}{N})^{0.5}$ and $cov_{m,m+1} = \frac{1}{N-1} \sum_{i=1}^{N} (x_{m,i} - \bar{x}_m)(x_{m+1,i} - \bar{x}_{m+1})$. $\bar{x}_m$ and $\bar{x}_{m+1}$ are the means in ROI at iteration m and $m+1$, var is the corresponding variances across the image elements. cov is the covariance across the two iterations. t is calculated until less than $t_{0.05}$ in the t-table to show a confidence level of 95% that the difference between images is small enough.

3   Experiment and Results

3.1   Monte Carlo Simulated Dynamic PET Data

Experiment Settings. First dataset is generated using Monte Carlo simulations of the acquisition of our PET scanner. There are totally 90 studies.

1. A series of 15 phantoms is generated based on Zubal thorax phantom of different sizes (to represent different patients from skinny to fat). Randomly generated motions (shifts and rotation) are added to each phantom.

Table 1. Kinetic parameters used for Monte Carlo simulations

ROI1				ROI2			
k1	k2	k3	k4	k1	k2	k3	k4
0.102	0.130	0.062	0.0068	0.082	0.102	0.045	0.0041
ROI3				Lesion			
k1	k2	k3	k4	k1	k2	k3	k4
0.064	0.124	0.042	0.0035	0.4870	0.7120	0.1950	0.0341
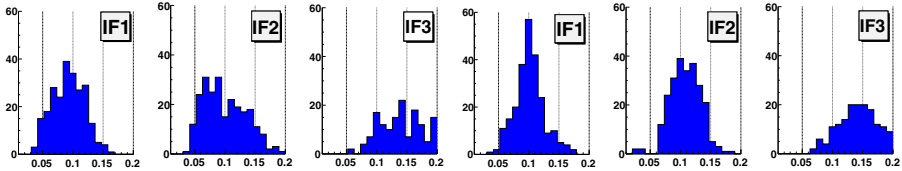
Fig. 3. Influx rates by General Method (left 3) and Our Method (right 3)

2. A lesion is added in every phantom. For each lesion, 2 cases are studied: true lesion (e.g. tumor) and false lesion (e.g. some normal tissue with undesired radiotracer concentration.). First 5 phantoms with lesions are shown in Fig. 1.

3. The Time Activity Curves (TAC) are generated by Feng input, $C_P^{FDG}(t) = (A_1 t - A_2 - A_3)e^{-\lambda_1 t} + A_2 e^{-\lambda_2 t} + A_3 e^{-\lambda_3 t}$. The values of the parameters λ_i and A_i selected here are $A_1 = 28\mu Ci/mL/min$, $A_2 = 0.75\mu Ci/mL$, $A_3 = 0.70\mu Ci/mL$, $\lambda_1 = 4.1339 min^{-1}$, $\lambda_2 = 0.01043 min^{-1}$ and $\lambda_3 = 0.1191 min^{-1}$. For Feng input, the final results need be calibrated for possible underestimation. The dynamic acquisition consists of 29 frames: $6 \times 5sec$, $2 \times 15sec$, $6 \times 0.5min$, $3 \times 2min$, $2 \times 5min$ and $10 \times 10min$. The kinetic parameters used in simulations are listed in Table. 1. The TACs of 3 normal ROIs are shown in Fig. 2 (b) and TACs of true lesion and false lesion are shown in Fig. 2 (c). False lesion uses the same kinetic parameters as ROI2.

4. All the parameter reconstructions are performed with 3 different input function initializations: Input Function 1 is perfect input function (equivalent to perfect blood sampling with less than 5% error), Input Function 2 is good input function (equivalent to a disturbed blood sampling with about 20% error), Input Function 3 is an Image Derived Input Function (IDIF). In next part, reconstructions with general purpose SM are called General Method.

Experiment Results. The influx rate maps are calculated based on **Influx Rate** $= \frac{k_1 k_3}{k_2 + k_3}$ to evaluate the reconstruction results.

First, all the true lesions are extracted from 45 studies and analyzed pixel-wisely. Fig. 3 shows the histograms of influx rates of lesions calculated by results from General Method and Our Method using different input functions. When using Input Function 1 and Input Function 2, our results are closer to the true value (it is 0.1047), and show obvious smaller standard derivations, which will help identify lesion sizes more precisely. With Input Function 3, the bad
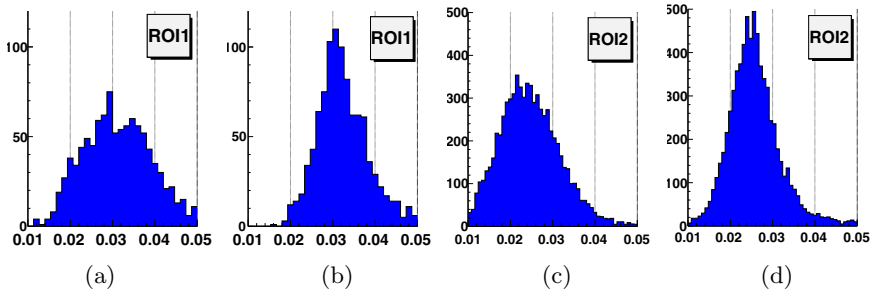
Fig. 4. Histograms of influx rate by (a) General Method in ROI1; (b) Our Method in ROI1; (c) General Method in ROI2; (d) Our Method in ROI2

estimation of input function leads to the overestimation of influx rates, however, our method still has more pixels near the true value. Our Method also introduces improvements in other regions. For ROI 1 and 2 indicated by Fig. 2 (a), in Fig. 4 we show the histograms of influx rates from all the pixels of all studies in corresponding ROIs. As the lesion region, results from our method are closer to the true values with obviously smaller standard derivations.

Fig. 5 shows the influx rate maps of Phantom 3 using Input Function 2. This special case is to represents the oversized patient with underestimations of the lesion. Our method first shows a better overall image quality. Then our method also show smoother results and better discrimination of lesion region and different ROIs like indicated by histograms in Fig. 3 and Fig. 4. The lesions are more clear and uniform by our method, and especially the false lesion is identified clearly to metabolize like muscles.

We summarize all 90 studies in Table. 2, which shows the ratio of successful identification of lesions using different input functions ("Success" means the difference between mean of lesion region and true value should be less than 40%, which is just the value to separate lesions from muscles in our experiments when lesions are near body surface or heart, and the standard deviations of lesion regions should be less than 0.67 to correctly identify the sizes of lesions). Both methods performs well by using Input Function 1 (perfect input function). However, with Input Function 2 (with disturbances), the accuracy of General Method decreases, but our method still provides good results. The simulation results show the improvement in identifying lesions by our method, and the reduction of requirement of accurate input function.

3.2   Real Patient Experiments

The real patient data in this study was a dynamic PET scan acquired from a 28-year-old, 75kg male volunteer using our PET scanner. 10 mCi ^{18}F-FDG was injected and a dynamic acquisition of the thoracic cavity started just after injection. The acquisition consists of 40 time frames: $20 \times 0.5min$, $15 \times 1min$, and $5 \times 2min$. All corrections are performed properly with the software provided by the scanner. The input function is estimated by the image-derived method. Fig. 6

(a)   (b)   (c)   (d)
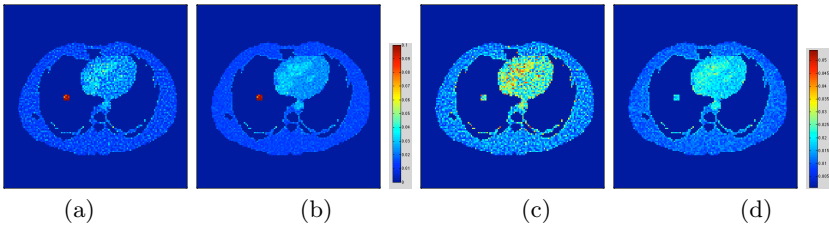
Fig. 5. Influx rate maps of Phantom 3. (a) General Method with true lesion; (b) Our Method with true lesion; (c) General Method with false lesion; (d) Our Method with false lesion.

Table 2. Summary of experiments

	Group1	Group2	Group3	Group4	Group5	Group6
Lesion Type	True	True	True	False	False	False
Input Function	1	2	3	1	2	3

Successful Estimation/Total Studies	Group1	Group2	Group3	Group4	Group5	Group6
General Mehtod	13/15	7/15	5/15	11/15	7/15	6/15
Our Mehtod	14/15	13/15	7/15	12/15	12/15	7/15

Muscles	0.0054
Our Method	0.0071
General Method	0.0085

(a)   (b)   (c)
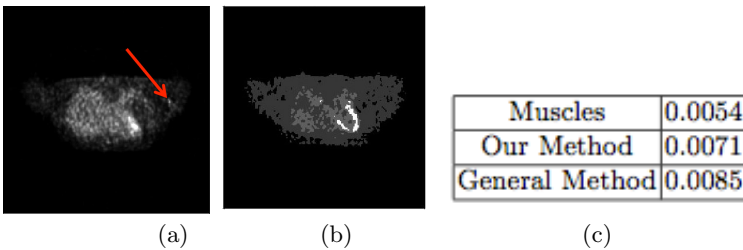
Fig. 6. (a) Lesion in 40th slice; (b) 32nd slice; (c) Influx rates

(a) shows a lesion region by a red arrow in the 40th slice. We calculate the influx rates of the lesion and compare them with the heart muscles in the 32nd slice. The lesion metabolism calculated by our method is closer to the muscles than that by General Method, and the lesion is confirmed by the doctor as a false lesion with temporarily increased metabolism than muscles, results from our method show potential improvement in diagnosis.

4 Conclusion

We presented a dynamic PET metabolism analysis framework by defining a patient adaptive SM. Experiment results show obvious improvements on identifying lesions by our method, and requirement of input functions is also reduced.

References

1. Rahmim, A., Tang, J., Zaidi, H.: Four-Dimensional (4D) Image Reconstruction Strategies in Dynamic PET: Beyond Conventional Independent Frame Reconstruction. Medical Physics 36(8), 3654 (2009)
2. Wang, G., Qi, J.: Generalized Algorithms for Direct Reconstruction of Parametric Images From Dynamic PET Data. IEEE Transactions on Medical Imaging 28(11), 1717–1726 (2009)
3. Gao, F., Liu, H., Shi, P.: Robust Estimation of Kinetic Parameters in Dynamic PET Imaging. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part I. LNCS, vol. 6891, pp. 492–499. Springer, Heidelberg (2011)
4. Moehrs, S., Defrise, M., Belcari, N., Guerra, A.D., Bartoli, A., Fabbri, S., Zanetti, G.: Multi-Ray-Based System Matrix Generation for 3D PET Reconstruction. Physics in Medicine and Biology 53(23), 6925 (2008)
5. Panin, V., Kehren, F., Rothfuss, H., Hu, D., Michel, C., Casey, M.: PET Reconstruction with System Matrix Derived from Point Source Measurements. IEEE Transactions on Nuclear Science 53(1), 152–159 (2006)
6. Zhang, L., Staelens, S., Holen, R.V., Beenhouwer, J.D., Verhaeghe, J., Kawrakow, I., Vandenberghe, S.: Fast and Memory-Efficient Monte Carlo-Based Image Reconstruction for Whole-Body PET. Medical Physics 37(7), 3667 (2010)
7. Su, K.H., Wu, L.C., Lee, J.S., Liu, R.S., Chen, J.C.: A Novel Method to Improve Image Quality for 2-D Small Animal PET Reconstruction by Correcting a Monte Carlo-Simulated System Matrix Using an Artificial Neural Network. IEEE Transactions on Nuclear Science 56(3), 704–714 (2009)
8. Wu, H., Pal, D.O., Sullivan, J., Tai, Y.C.: A Feasibility Study of a Prototype PET Insert Device to Convert a General-Purpose Animal PET Scanner to Higher Resolution. Journal of Nuclear Medicine 49(1), 79–87 (2008)
9. Li, Z., Li, Q., Yu, X., Conti, P., Leahy, R.: Lesion Detection in Dynamic FDG-PET Using Matched Subspace Detection. IEEE Transactions on Medical Imaging 28(2), 230–240 (2009)
10. Laffon, E., de Clermont, H., Vernejoux, J.M., Jougon, J., Marthan, R.: Feasibility of Assessing [18F]FDG Lung Metabolism with Late Dynamic PET Imaging. Molecular Imaging and Biology 13(2), 378–384 (2011)
11. Vriens, D., Visser, E.P., de Geus-Oei, L.F., Oyen, W.J.G.: Methodological Considerations in Quantification of Oncological FDG PET Studies. European Journal of Nuclear Medicine and Molecular Imaging 37(7), 1408–1425 (2010)
12. Fessler, J., Erdogan, H.: A Paraboloidal Surrogates Algorithm for Convergent Penalized-Likelihood Emission Image Reconstruction. In: Nuclear Science Symposium, Conference Record, vol. 2, pp. 1132–1135 (1998)

A Personalized Biomechanical Model for Respiratory Motion Prediction

B. Fuerst[1,2], T. Mansi[1], Jianwen Zhang[1], P. Khurd[1], J. Declerck[3],
T. Boettger[4], Nassir Navab[2], J. Bayouth[5], Dorin Comaniciu[1], and A. Kamen[1]

[1] Siemens Corporation, Corporate Research and Technology, Princeton, NJ, USA
[2] Computer Aided Medical Procedures, Technische Universität München, Germany
[3] Siemens Molecular Imaging, Oxford, United Kingdom
[4] Oncology Care Systems, Siemens AG, Heidelberg, Germany
[5] Iowa University, Iowa City, IA, USA

Abstract. Time-resolved imaging of the thorax or abdominal area is affected by respiratory motion. Nowadays, one-dimensional respiratory surrogates are used to estimate the current state of the lung during its cycle, but with rather poor results. This paper presents a framework to predict the 3D lung motion based on a patient-specific finite element model of respiratory mechanics estimated from two CT images at end of inspiration (EI) and end of expiration (EE). We first segment the lung, thorax and sub-diaphragm organs automatically using a machine-learning algorithm. Then, a biomechanical model of the lung, thorax and sub-diaphragm is employed to compute the 3D respiratory motion. Our model is driven by thoracic pressures, estimated automatically from the EE and EI images using a trust-region approach. Finally, lung motion is predicted by modulating the thoracic pressures. The effectiveness of our approach is evaluated by predicting lung deformation during exhale on five DIR-Lab datasets. Several personalization strategies are tested, showing that an average error of $3.88 \pm 1.54\,mm$ in predicted landmark positions can be achieved. Since our approach is generative, it may constitute a 3D surrogate information for more accurate medical image reconstruction and patient respiratory analysis.

1 Introduction

Respiratory motion is a source of artifacts in medical image acquisition, which is the basis for disease monitoring, therapy planning and intervention guidance. Currently, signals from devices such as spirometers, abdominal pressure belts or external markers are used as surrogates of respiratory motion. However, the one-dimensional nature of these signals makes it difficult to estimate the 3D lung deformation accurately. There is therefore a need for methods to predict the 3D lung deformation during regular and irregular breathing cycles.

Two categories of methods for patient-specific estimation of respiratory movements can be distinguished: image-based and biomechanical methods. On the one hand, image-based methods commonly estimate the lung deformations between

two or more phases of the respiratory cycle using non linear image registration. The idea is to create a lung-motion atlas, which is subsequently adapted to the patient [4]. Advanced image registration techniques have been developed to that end, relying in particular on advanced spatial regularization terms that allow sliding [8,10]. Recent results are promising, but image-based methods are usually restricted to normal breathing patterns [13], with lower predictive power and versatility in patients. On the other hand, biomechanical approaches compute lung biomechanics, often using finite element methods (FEM), to simulate the physiological deformations during respiration cycles [1,12]. The anatomical model is created from patient images and then deformed according to lung tissue properties and non-physiological driving forces defined from the difference between two volumetric objects [1,12]. As a result, these approaches still rely on 4D image data, which makes it difficult to predict respiratory motion when unexpected changes in breathing patterns appear [9,11].

As a first step towards the prediction of respiratory motion given 1D surrogate signals of the thorax displacements, we propose a generative biomechanical model of the respiratory system driven by patient-specific thoracic and diaphragmatic pressures (Sec. 2). Contrary to previous approaches, our framework is not directly driven by image forces but by a novel thorax/diaphragm/lung interaction model. Deformation is not limited by a secondary geometry or based on image data. As illustrated in Fig. 1, the framework first estimates a comprehensive anatomical model from an image at end of exhale (EE). If not available, the thoracic and diaphragmatic pressures necessary to load the lung till end of inhale (EI) are estimated automatically using a trust region optimizer. Lung deformation is then predicted throughout the respiratory cycle by modulating the thoracic pressures. The prediction power of our model is evaluated by predicting exhale deformations in five DIR-Lab datasets (Sec. 3). We further investigate the need of including information about the deformation into personalization and show in experiments that an average error of $3.88 \pm 1.54\,mm$ in predicted landmark positions can be achieved. Sec. 4 concludes the paper.
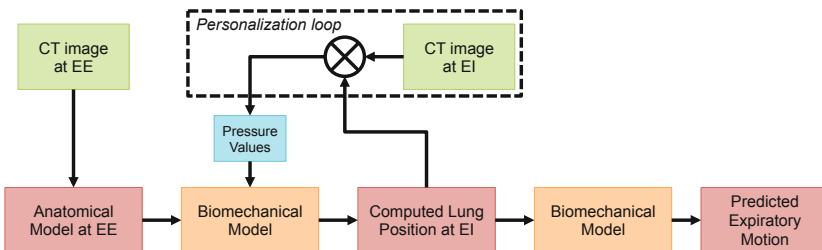
Fig. 1. Pipeline of the proposed motion model. See text for details

2   Methods

The motion model consists of two main components: *i)* a detailed anatomical model of the respiratory system comprising the lungs, thorax, and a sub-diaphragm region grouping abdominal organs including the diaphragm (Sec. 2.1); and *ii)* a generative biomechanical model that computes lung deformations according to tissue properties and pressures generated by the thorax and the diaphragm (Sec. 2.2). The pressure parameters of the model are estimated automatically from two EE and EI CT images (Sec. 2.3).

2.1   Anatomical Model of Respiratory System

Thorax dynamics are complex, involving several interacting organs. In particular, the lung is deformed indirectly through expansion and contraction of the rib cage and diaphragm. During this process, the lung slides along the thoracic surface of the rib cage and diaphragm. The diaphragm is fixed along the rib cage and is curved towards the lung, and in some patients it slides along the rib cage as well. To capture these interactions, we create an anatomical model that consists of four independent components: the two lungs, the thorax, and a lumped component, called sub-diaphragm region, which includes the diaphragm and other abdominal organs. The two lungs and patient's skin, which defines the outer layer of the thorax, are automatically segmented from 3D CT images using a machine learning approach combined with level-set optimization, as described in [5] (Fig. 2, left panel). The sub-diaphragm region is computed automatically from the skin and the lung segmentations by casting the lung downwards (Fig. 2).

Each component is meshed with tetrahedra using CGAL (*www.cgal.org*). To capture the heterogeneous muscle forces, thoracic and diaphragmatic pressures are estimated regionally. The inner surface of the thorax and the diaphragmatic interface are sub-divided into uniform patches (Fig. 2 right panel). The number of patches is set experimentally; in areas with lung interaction 9 evenly distributed thoracic patches (incl. mediastinum), and two on the diaphragm.

2.2   Biomechanical Model of Respiratory System

During respiration, the muscles apply forces, which are then transferred to the lungs. We thus deform the lung by solving simultaneously the dynamics equations for the lungs (l), thorax (t) and sub-diaphragm region (d):

$$\begin{cases} \mathrm{M}^l\ddot{\mathbf{U}}^l + \mathrm{C}^l\dot{\mathbf{U}}^l + \mathrm{K}^l\mathbf{U}^l = \mathbf{F}_c^{t\to l} + \mathbf{F}_c^{d\to l} \\ \mathrm{M}^t\ddot{\mathbf{U}}^t + \mathrm{C}^t\dot{\mathbf{U}}^t + \mathrm{K}^t\mathbf{U}^t = \mathbf{F}_c^{l\to t} + \mathbf{F}_c^{d\to t} + \mathbf{F}_p^t \\ \mathrm{M}^d\ddot{\mathbf{U}}^d + \mathrm{C}^d\dot{\mathbf{U}}^d + \mathrm{K}^d\mathbf{U}^d = \mathbf{F}_c^{l\to d} + \mathbf{F}_c^{t\to d} + \mathbf{F}_p^d \end{cases} \qquad (1)$$

In the following, the superscripts are omitted if not necessary. The vectors $\ddot{\mathbf{U}}$, $\dot{\mathbf{U}}$ and $\mathbf{U}$ gather the accelerations, velocities and positions of the free nodes of
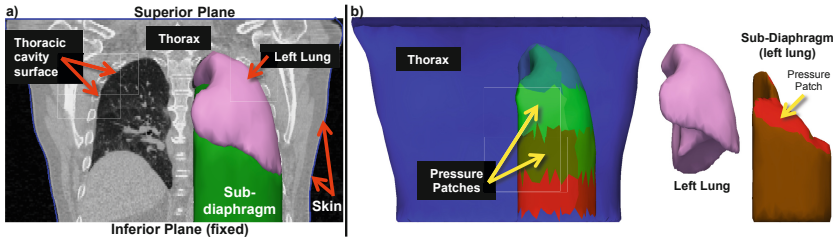
Fig. 2. *a)*: The coronal CT slice is overlaid by the skin segmentation (blue) and intersects in 3D with the left lung (pink) and the sub-diaphragm mesh (green). *b)*: The thorax (front face removed) with different coloring for each pressure patch for the left lung. Below the lung lays the sub-diaphragm mesh (brown; one pressure patch visible in red) comprising the diaphragm and the abdominal organs. For more realistic deformations, this mesh is separated from the thorax.

each simulated object. M is the (lumped) mass matrix (mass densities: $\rho^l = 1.05\,g/mL$, $\rho^{\{t,d\}} = 1.50\,g/mL$). K is the stiffness matrix of the internal elastic forces. C is a damping matrix. Here, Rayleigh damping is chosen, with coefficients 0.1 for both mass and stiffness. As described below, lungs, thorax and sub-diaphragm are each subject to interaction forces F_c, which model the sliding interactions between two organs, and pressure forces F_p, which represents the physiological force driving the respiratory motion.

Passive Material Properties. Lungs, thorax and diaphragm are non linear, heterogeneous materials [9,11]. Because we want to estimate the driving pressures and lung stiffness, fast simulations are necessary to reduce computation time. We thus chose to use a linear elastic model [1], whose stiffness is set by the Young's modulus E and compressibility by Poisson's ratio ν. Co-rotational linear tetrahedra are employed to cope with large deformations [6]. The parameters are set using values reported in the literature [1]: $E^l = 900\,Pa$, $E^{\{t,d\}} = 7800\,Pa$, $\nu^l = 0.4$ and $\nu^{\{t,d\}} = 0.43$.

Respiratory Forces. Breathing relies on the contraction of surrounding muscles, which expand the lungs during contraction. In our model, we apply the negative pressure as force on each element of the thoracic surface of the thorax and diaphragm: $\mathbf{f}_p^i = p^i\,\mathbf{n}\,dS$, where p^i is the pressure of the i-th patch and $\mathbf{n}$ is the normal of the surface element dS.

Collision and Sliding Interaction. Between lung and thorax lies the pleura, which is filled with a serous fluid allowing nearly friction-free movement between the lungs and thoracic cavity. During respiration, the lungs stay attached with the thoracic surface of the rib cage and diaphragm, the change of pleural volume being minimal. To model this behavior, a collision model has been implemented as a penalty force to prevent interpenetration, allow tangential sliding movement, and attract objects to each other to ensure they stay connected. The collision detection, based on proximity detection, is provided by the SOFA framework[1].

---

[1] www.sofa-framework.org
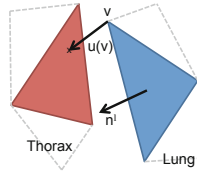
Fig. 3. Drawing of the collision force. By design, that force does not restrict sliding but only penalizes inter-penetration of the thorax and lung.

Once a collision is detected, the proposed penalty force is applied automatically whenever the distance between a vertex of one object and a triangle of another object is lower than an alarm distance d_a. Exemplarily for all interactions between lungs, thorax and diaphragm the force generated by the thorax on a vertex **v** of the lung can be written as (Fig. 3):

$$\begin{cases} \mathbf{F}_c^{t \to l}(\mathbf{v}) = \mathbf{0}, \text{ if } \|\mathbf{u}(\mathbf{v})\| = 0 \text{ or } \|\mathbf{u}(\mathbf{v})\| > d_a \\ \mathbf{F}_c^{t \to l}(\mathbf{v}) = -\mathbf{n}^l k_s \left(\mathbf{u}(\mathbf{v}) \cdot \mathbf{n}^l \right) \text{ otherwise} \end{cases}$$

where **u** is the vector between the lung vertex **v**, which belongs to the triangle $\mathcal{T}^l$, and the corresponding collision point on the thorax. $\mathbf{n}^l$ is the normal of the triangle $\mathcal{T}^l$. k_s is the penalty force stiffness coefficient, set to $0.1 N/m$ in this study. The interactions F_c between all three objects are defined in a similar way. *Implementation* The biomechanical model is implemented based on SOFA framework[1]. Eq. 1 is solved using a semi-implicit Euler solver with a time step of $1ms$.

2.3   Model Personalization

Model personalization is achieved by optimizing the patch-wise pressure values. We estimate the pressure necessary to load the lung from EE to EI by minimizing a multi variate cost function using Powell's NEWUOA algorithm [7], a trust-region method that does not explicitly calculate cost function gradients.

The cost function is defined by $\mathcal{C} = \mathcal{D}_S + \mathcal{D}_{LM}$, where $\mathcal{D}_S$ is the mean Hausdorff surface-to-surface distance between the deformed EE lung surface at system equilibrium and the segmented lung surface at EI. $\mathcal{D}_{LM}$ is the average Euclidian distance between landmarks at EI and their corresponding EE landmarks moved according to the internal deformation provided by the biomechanical models.

3   Experiments and Results

Two different sets of experiments were carried out to evaluate our approach and assess the importance of considering internal deformation during personalization. Tab. 1 summarizes the different protocols. All configurations were validated by predicting full exhale (EI to EE), without any image information.

Table 1. Experiment protocols. See text for details.

Personalization (Optimization Parameters)	Cost Function
Experiment 1 (E1): Personalization based on surface distance	
14 pressure values	$\mathcal{C} = \mathcal{D}_S$
Experiment 2 (E2): Personalization based on surface and landmark distances	
14 pressure values	$\mathcal{C} = \mathcal{D}_S + \mathcal{D}_{LM}$

Phase 1 (EI)   2   3   4   5   6 (EE)
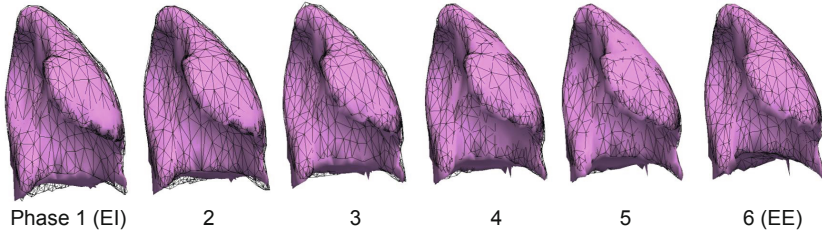
Fig. 4. A simulated lung during exhale (pink). The ground truth (in black wireframe) and the simulation (in pink) are showed for each 4D CT phase.

Data Sets and Pre-Processing. We used EI and EE images from DIR Lab data sets [2] for which the thorax is entirely visible (cases 6 to 10, image resolution of $0.97 \times 0.97 \times 2.50\,mm$). For these sets landmarks were available. We automatically segmented the lungs and skin surface and meshed the thorax, lung, and diaphragm including abdominal organs with 25351, 2650 and 2754 tetrahedra on average. In the following we report results for the left lung only, but due to the anatomical separation, our method can easily be extended to both lungs.

Personalization. The pressure estimation was based on a cost function (see Tab. 1), which compared the current state of the simulation with the ground truth at EI. The NEWUOA optimizer converged after an average of 109 iterations for pressure estimation ($\approx$ 15h single threaded system with 2.93GHz and 6GB memory; improvements expected by use of multi threading and collision detection on GPU, such as in [3]). The obtained landmark errors at EI (Tab. 2) were of the same order of magnitude as values reported in the literature [10,12], although our model is not directly driven by image forces.

Table 2. Mean errors between simulation and ground truth. Setup details in Tab 1.

Case	6	7	8	9	10	Mean
E1: $\mathcal{C} = \mathcal{D}_S$						
EI surface error (mm)	4.66	3.73	4.43	3.09	2.95	3.77 ± 0.89
EI landmark error (mm)	4.68	7.32	11.17	5.25	2.85	6.26 ± 4.92
Mean landmark error during prediction (mm)	3.96	4.67	6.35	3.29	2.83	4.22 ± 2.13
E2: $\mathcal{C} = \mathcal{D}_S + \mathcal{D}_{LM}$						
EI surface error (mm)	4.08	3.66	4.35	3.17	3.04	3.66 ± 0.69
EI landmark error (mm)	4.53	6.86	8.63	4.53	2.32	5.37 ± 3.26
Mean landmark error during prediction (mm)	3.67	4.55	5.41	3.18	2.56	3.88 ± 1.54
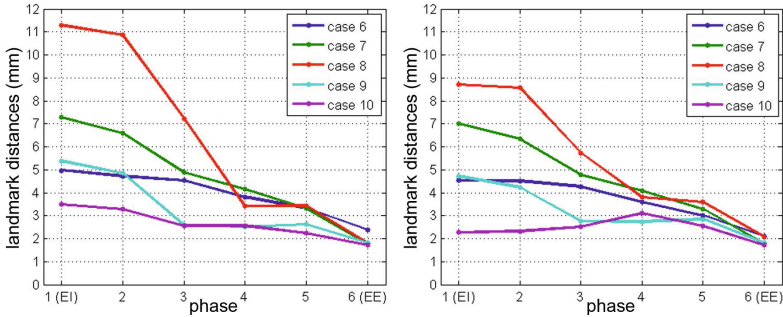
Fig. 5. Average landmark error during exhale simulation for the two experiment scenarios: estimation of patch-wise pressure values using the surface distance (left) or the surface plus landmark distance (right) as cost function. The simulation starts at the mechanical no-load phase EE, the lung is loaded using the personalized pressure, and recoils to the no-load phase (plotted exhale). Therefore the errors near EE are smaller.

Validation. The quality of the model's exhale prediction was evaluated by comparing the simulated landmark positions with the landmarks of each intermediate phase during exhale, which were not used during personalization. The exhale was performed by turning off the personalized thoracic pressures just after reaching the EI equilibrium (Fig 4). To reproduce a real-case scenario we synchronized the simulated lung with the 4D CT images by means of the lung volume. Despite the simplifications of our model, we obtained promising predicted lung motion, with an average landmark error of $3.88 \pm 1.54\,mm$ (Tab. 2, Fig. 5).

The experiments showed that the predictive power of the model can be improved by considering the landmarks as information of internal deformation into the cost function; surface matching is not enough. Analysis based on the distances of each landmark using a paired t-test showed that the improvements of the model's exhale prediction obtained in E2 were significant for cases 6 to 9 (p-value < 0.05). Error in case 10 was already of the order of magnitude of the slice thickness, therefore a significant improvement was not expected.

4   Discussion and Future Works

In this study we have presented a novel approach to predict patient respiratory motion, including the estimation of the thoracic pressure values. To the best of our knowledge, our generative model is the first to simulate the lung, thorax and diaphragm interactions without being explicitly driven by image forces. Personalized from two CT images, the model is generative and can predict respiratory motion throughout the entire cycle. The obtained results were of the same order of magnitude as state-of-the-art respiratory motion models, encouraging further work in this direction. Tethering the lungs to the airways, estimating spatial varying tissue properties, using hyperelastic material, refining the anatomical model, and reducing the uncertainties in the FEM simulation could reduce the simulation error. In a step towards the clinical application, we plan to investigate

the correlation between the 1D surrogate and personalized pressure force field. In conclusion, our method, being generative, may constitute a 4D surrogate model to improve prediction of respiratory motion for image reconstruction.

References

1. Al-Mayah, A., Moseley, J., Velec, M., Brock, K.: Sliding characteristic and material compressibility of human lung: Parametric study and verification. Medical Physics 36, 4625–4633 (2009)
2. Castillo, R., Castillo, E., Guerra, R., Johnson, V., McPhail, T., Garg, A., Guerrero, T.: A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets. Physics in Medicine and Biology 54, 1849–1870 (2009)
3. Courtecuisse, H., Cotin, S., Allard, J., Soler, L.: GPU-based interactive simulation of liver resection. In: ACM SIGGRAPH 2011 Computer Animation Festival, p. 98. ACM (2011)
4. Ehrhardt, J., Werner, R., Schmidt-Richberg, A., Schulz, B., Handels, H.: Generation of a mean motion model of the lung using 4D-CT image data. In: Eurographics Workshop on Visual Computing for Biomedicine, pp. 69–76 (2008)
5. Kohlberger, T., Sofka, M., Zhang, J., Birkbeck, N., Wetzl, J., Kaftan, J., Declerck, J., Zhou, S.K.: Automatic Multi-organ Segmentation Using Learning-Based Segmentation and Level Set Optimization. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part III. LNCS, vol. 6893, pp. 338–345. Springer, Heidelberg (2011)
6. Nesme, M., Payan, Y., Faure, F., et al.: Efficient, physically plausible finite elements. In: Eurographics 2005 Short Papers (2005)
7. Powell, M.: Developments of NEWUOA for minimization without derivatives. IMA Journal of Numerical Analysis 28(4), 649–664 (2008)
8. Rietzel, E., Chen, G., Choi, N., Willet, C.: Four-dimensional image-based treatment planning: Target volume segmentation and dose calculation in the presence of respiratory motion. International Journal of Radiation Oncology* Biology* Physics 61(5), 1535–1550 (2005)
9. Risholm, P., Ross, J., Washko, G.R., Wells, W.M.: Probabilistic Elastography: Estimating Lung Elasticity. In: Székely, G., Hahn, H.K. (eds.) IPMI 2011. LNCS, vol. 6801, pp. 699–710. Springer, Heidelberg (2011)
10. Schmidt-Richberg, A., Werner, R., Handels, H., Ehrhardt, J.: Estimation of Slipping Organ Motion by Registration with Direction-Dependent Regularization. Medical Image Analysis 16, 150–159 (2011)
11. Villard, P.F., Beuve, M., Shariat, B., Baudet, V., Jaillet, F.: Simulation of lung behaviour with finite elements: Influence of bio-mechanical parameters. In: Medical Information Visualisation-Biomedical Visualisation 2005, pp. 9–14. IEEE (2005)
12. Werner, R., Ehrhardt, J., Schmidt, R., Handels, H.: Patient-specific finite element modeling of respiratory lung motion using 4D CT image data. Medical Physics 36, 1500–1511 (2009)
13. Zhang, T., Jeraj, R., Keller, H., Lu, W., Olivera, G., Mackie, T., Paliwal, B.: Treatment plan optimization incorporating respiratory motion. Medical Physics 31, 1576–1586 (2004)

Endoscope Distortion Correction Does Not (Easily) Improve Mucosa-Based Classification of Celiac Disease[⋆]

Jutta Hämmerle-Uhl[1], Yvonne Höller[1], Andreas Uhl[1], and Andreas Vécsei[2]

[1] Department of Computer Sciences
University of Salzburg, Austria
[2] St.Anna Children's Hospital, Dept. Pediatrics
Medical University, Vienna
andreas.uhl@sbg.ac.at

Abstract. Distortion correction is applied to endoscopic duodenal imagery to improve automated classification of celiac disease affected mucosa patches. In a set of six edge- and shape-related feature extraction techniques, only a single one is able to consistently benefit from distortion correction, while for others, even a decrease of classification accuracy is observed. Different types of distortion correction do not lead to significantly different behaviour in the observed application scenario.

Keywords: endoscope distortion correction, shape- and edge-based features, celiac disease, automated classification.

1 Introduction

Computer-aided decision support systems relying on automated analysis of endoscopic imagery receive increasing attention [1].

A specific type of degradation, present in all endoscopic images, is a barrel-type distortion. This type of degradation is caused by the wide-angle (fish eye) nature of the optics used in endoscopes.

The aim of correcting this distortion in endoscopy is manifold. Barrel type distortion is claimed to affect diagnosis [2], since it introduces nonlinear changes in the image, due to which the outer areas of the image look significantly smaller than their actual size. Therefore, the estimation of area or perimeter of observed lesions can be significantly incorrect depending on the position in the image [3]. In a recent study [4] it has been demonstrated, that in classification of celiac disease based on duodenal images, in fact misclassification cases can be related to the extent of barrel distortion of the texture patches involved in classification. Using the same image material, the impact of distortion correction on classification accuracy has been investigated [5], [4]. In these studies it turned out that most feature extraction methods considered failed to take advantage of applying distortion correction as a pre-processing step to the endoscopic images, resulting in an even decreased classification accuracy. It has been suspected that these unexpected results might be due to the (i) (too) simple distortion correction technique applied. The only feature extraction techniques exhibiting improved classification

---

when applied to distortion corrected images were based on edges and geometrical features [5]. Therefore, it was also speculated that in general, (ii) edge and shape-related feature types would be able to benefit from distortion correction.

In this work we focus on those two conjectures (i) and (ii) stated as conclusions after result analysis in the mentioned studies. First, we employ the more recent parameter-free distortion correction approach of Hartley and Kang [6]. Second, we use a set of features related to edge and shape information instead of the mostly texture-oriented descriptors in [5,4]. Further contributions of this work are the usage of a more realistic evaluation protocol for classification assessment (leave-one-patient-out (LOPO) cross validation) and the application of a richer set of classifiers to avoid bias due to the use of a single classifier.

The manuscript is structured as follows. Section 2 explains the background of applying duodenal mucosa texture classification for diagnosis and staging of celiac disease and describes the image database used with the corresponding histological ground truth. In Section 3, we describe the experimental setup by first explaining the distortion correction techniques and their respective application to our image test and subsequently, by reviewing the feature extraction (and classification) techniques employed. Section 4 presents and discusses experimental results and in Section 5 we finally conclude this work.

2   Classification of Duodenal Texture for Celiac Disease Diagnosis

Celiac disease, commonly known as gluten intolerance, is a complex autoimmune disorder that affects the small bowel in genetically predisposed individuals of all age groups after introduction of food containing gluten. Endoscopy with biopsy is currently considered the gold standard for the diagnosis of celiac disease. During standard upper endoscopy at least four duodenal biopsies are taken. Microscopic changes within these specimen are then classified in a histological analysis according to the Marsh classification. The modified Marsh classification [7] distinguishes between classes Marsh-0 to Marsh-3, with subclasses Marsh-3a, Marsh-3b, and Marsh-3c, resulting in a total number of six classes. An automated system identifying areas affected by celiac disease in the duodenum can help to improve biopsy reliability (by indicating areas eventually affected by celiac disease), can aid to improve less invasive diagnosis techniques avoiding biopsies, and can reduce the costs of interpreting video material captured during capsule endoscopy [7]. Prior approaches dealing with the computer-aided diagnosis of celiac disease using endoscopic still images images include feature extraction based on Local Binary Pattern based operators, band-pass type Fourier filters, histogram and wavelet-transform based features, as well as smoothness/sharpness measures [7]. Techniques involving temporal information computed from video-capsule endoscopy have been also described [8].

The image test set used in this work (see [7] for example images) stems from three pediatric gastroscopes without magnification, types GIF-Q165 and GIF-N180, Olympus, with two of the first type and one of the latter type, respectively. The patients presented in the pediatric Department because of celiac-like symptoms. Diagnostic evaluation was indicated because of dyspeptic symptoms, positive celiac serology,

anaemia, malabsorption syndromes, inflammatory bowel disease, and gastrointestinal bleeding. For the endoscopy, the modified immersion technique was applied. This procedure is based on the instillation of water into the duodenal lumen for better visibility of the villi. Then, the tip of the gastroscope is inserted into the water in order to take images of meaningful regions. The images were taken from the Duodenal Bulb and the Pars Descendens. Most importantly, these regions differ by their geometric properties. Thus, it is necessary to treat these image sets separately.

From the acquired images, an experienced endoscopist extracted 128×128 pixels patches significant for diagnosis. The images and patients were pre-classified by the diagnostic outcome of the biopsy of the significant region at the hospital into the modified Marsh classification as shown in Table 1.

Table 1. Number of images/patients in the data sets and Marsh-classes

Data-Set	Marsh-0	Marsh-3a	Marsh-3b	Marsh-3c
Bulbus	163/60	47/8	54/8	23/8
Pars Descencens	141/72	47/10	60/8	72/12

3   Experimental Study

3.1   Distortion Correction

The following stages are applied to the entire endoscopic images before the extraction of the texture patches used for classification. Each colour image has been transformed into a grayscale image with the usual conversion formula and subsequently, the MATLAB built-in function for histogram equalisation (flat histogram) has been applied.

We use a planar checkerboard pattern (with points on a known grid) for distortion calibration (see Fig. 1.a). Fig. 1.b shows an example of a distortion corrected calibration pattern. The first distortion correction technique applies the MATLAB software developed by J.-Y. Bouguet[1]. For each gastroscope, 10 images were chosen to extract calibration points and distorted points. The algorithm to extract the grid corners of a checkerboard requires clicking on the four extreme corners of the rectangular checkerboard pattern. The calibration images that were used in this study contain only parts of a checkerboard, so that for each image some sensible area of the checkerboard was decided to contain the most extreme corners. Since the distortion in this study is quite significant, providing a manual estimation for radial distortion was required for all images. After computing the intrinsic and extrinsic camera parameters, the undistort-tool as provided in the tool-box is used (which is also used for computing the centres of the texture patches as mentioned below).

Barreto et al. [9] have found the parameter-free approach of Hartley and Kang [6] being better suited for endoscopic imagery as compared to Bouguet's approach – therefore, we have developed a corresponding MATLAB implementation of their technique. After a manual extraction of calibration points and their adjustment using Bouguet's

---

[1] http://www.vision.caltech.edu/bouguetj/calib_doc/
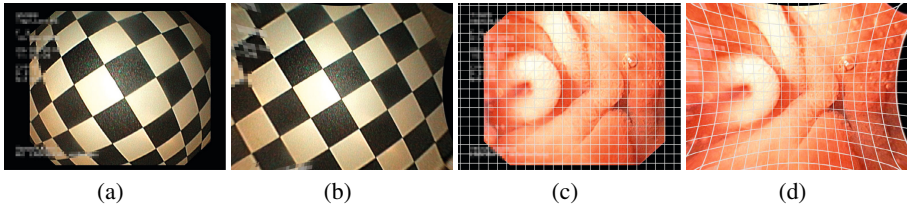
Fig. 1. Distortion correction applied to checkerboard (taken with Olympus GIF-Q165) and to entire endoscopic image (Bouguet distortion correction applied)

tool, the implementation of Peter Kovesi[2] was applied to all images to calculate the fundamental matrix. The reminder of the algorithm is implemented as described in the paper, partially using MATLAB built-in functions for e.g. optimisation ("fmincon").

Since after distortion correction the squared texture patches using for classification do no longer correspond to squares (see Fig. 1.c and 1.d) these cannot be used immediately for subsequent classification (most feature extraction techniques implicitly assume at least a rectangularly shaped texture patch). Therefore we apply the following technique to generate square-shaped texture from distortion corrected image material: Based on the original (distorted) endoscopic images, we record the coordinates of the centre of the extracted 128×128 pixels. Subsequently, distortion correction is applied to the entire original images and the recorded centre coordinates are mapped into the distortion corrected image. Using these coordinates, a 128×128 pixels texture square is extracted from the distortion corrected image which is then used for classification.

3.2   Feature Extraction and Classification

To be able to assess the impact of distortion correction techniques on the classification accuracy, we use a set of different feature extraction techniques. Contrasting to earlier studies, emphasis is given to edge- and shape-related strategies.

Fractal Dimension: Boxcounting [10]: A texture signature is computed from binary images obtained from original images using different thresholds and application of the box-counting fractal dimension on each thresholded image. In our implementation, all gray-level thresholds from 50 to 175 were used to generate binary images, for box-counting, Moisy's tool[3] is used with box-sizes from 2^2 - 2^{32}. For the final signature, the mean and standard deviation over the values for the box sizes are used for each threshold value.

Locally Invariant Fractal Features [11]: Local fractal dimension (also termed local density) is computed for each pixel in an image after applying the MR8 filterbank. For each class, the 8-dimensional local density vectors of all training images are aggregated and subjected to k-means clustering resulting in cluster centres termed textons. For an image to be classified, local density vectors are computed and each one is labelled with

---

[2] www.csse.uwa.edu.au/~pk/research/matlabfns/
[3] www.fast.u-psud.fr/~moisy/ml/

the texton that is closest to it. The frequency histograms of the texton occurrences are used as feature descriptors. Custom MATLAB code is developed for this approach.

Gray-Level Co-occurrence Matrix (GLCM [12]): The GLCM is defined over an image as the distribution of co-occurring values at a given offset Δx and Δy for a $n \times m$ image as follows:

$$C_{\Delta x, \Delta y}(i,j) = \sum_{p=1}^{n} \sum_{q=1}^{m} \begin{cases} 1, & if \ I(p,q) = i \ and \ I(p+\Delta x, q+\Delta y) = j \\ 0, & otherwise \end{cases} \quad (1)$$

For classification, the Haralick features contrast, correlation, energy, and homogeneity are used for offset-values 1,2,4, and 8 in 4 directions (vertical, horizontal, diagonal 45° and 135°). Calculations are performed with the MATLAB built-in function "graycomatrix".

Edge Co-occurrence Matrix (ECM [13]): For this approach (custom MATLAB implementation), a Sobel edge detection approach using the Robinson compass masks is applied to the images. Subsequently, a co-occurrence matrix is constructed using 1,2,3, and 4 as distances in 8 directions. Again, the Haralick features computed from this matrix as mentioned above are used as feature vector.

Edge Orientation Histogram (EH [14]): The EH is one of three MPEG-7 texture descriptors. The EH requires dividing the image into 4×4 sub-images, where each sub-image is sub-divided again into blocks of typically 4×4 pixels. The EH finds vertical, horizontal, diagonal and non-directional edges. This makes the EH specifically well suited for natural images with non-uniform edge distribution. Each image block is filtered to obtain the most prominent edge in the block. If the block is monotone, no edge is counted. As a consequence, a histogram with 5 bins can be computed over all the image blocks in each of the 16 sub-images. Thus, this results in an 80 bin histogram which is computed using the implementation of OConaire[4].

Spatial Size Distributions (SSD [15]): The difference of the autocorrelation for a given image and the autocorrelation of the same image after applying a morphological opening is computed. To be able to capture texture properties of different sizes, these computations are performed using scaled versions of the structure element used. The results obtained for the different structure element sizes are then summed up and normalised by the square of the sum over all grayscale values within the image. The result is a cumulative distribution function. The probability density associated with this cumulative distribution is called a spatial size distribution. The features which are then used to classify textures are obtained by computing first-order and second-order moments of the probability density.

While the shape variants of the structuring element have been chosen in accordance to the original suggestion, the number of scales of the structuring element and the

---

[4] `clickdamage.com/sourcecode/index.php`

number of different support disc sizes have been reduced to 4 and 2 in our MATLAB implementation, respectively, to limit the very demanding computations.

Classification: To avoid over-fitting phenomena the leave-one-patient-out cross-validation protocol is used to estimate classification accuracy. In each validation run, MATLAB built in classifiers are used: Discriminant analysis with a diagonal quadratic function, knn classification (with Euclidean distance metric and $k = 1$), and classification with SVM using a linear or a quadratic kernel function, respectively. Classification is performed separately for the two topographical regions of the duodenum. From the resulting classes that are assigned to the images, the performance of the classification was evaluated by calculating sensitivity, specificity and overall accuracy.

3.3 Experimental Results

In Tables 2 - 4 we display classification results. Configurations where results obtained from distortion corrected material are superior in terms of overall accuracy are marked in bold face.

Table 2. Classification performance for fractal dimension-based feature extraction

Fractal dimension: Boxcounting							**Locally invariant fractal features**					
class.	region	images	Sens.	Spec.	Acc.		class.	region	images	Sens.	Spec.	Acc.
diag.	**Bulbus**	**distorted**	.51	.66	.59		diag.	**Bulbus**	**distorted**	.35	.91	.67
quad.		**Hartley**	.58	.82	**.72**		quad.		**Hartley**	.30	.93	.66
		Bouguet	.46	.86	**.69**				**Bouguet**	.19	.96	.62
	Pars Desc	**distorted**	.46	.52	.49			**Pars Desc**	**distorted**	.64	.72	.68
		Hartley	.51	.65	**.57**				**Hartley**	.65	.64	.65
		Bouguet	.66	.57	**.62**				**Bouguet**	.47	.70	.57
knn	**Bulbus**	**distorted**	.40	.75	.60		knn	**Bulbus**	**distorted**	.73	.65	.69
		Hartley	.48	.79	**.66**				**Hartley**	.71	.54	.61
		Bouguet	.85	.65	**.65**				**Bouguet**	.56	.82	**.71**
	Pars Desc	**distorted**	.47	.57	.52			**Pars Desc**	**distorted**	.73	.51	.63
		Hartley	.50	.57	**.53**				**Hartley**	.56	.58	.57
		Bouguet	.45	.59	.51				**Bouguet**	.55	.51	.53
SVM-	**Bulbus**	**distorted**	.38	.72	.57		SVM	**Bulbus**	**distorted**	.73	.90	.83
linear		**Hartley**	.52	.79	**.67**		linear		**Hartley**	.59	.85	.74
		Bouguet	.53	.82	**.69**				**Bouguet**	.70	.88	.80
	Pars Desc	**distorted**	.57	.43	.51			**Pars Desc**	**distorted**	.73	.68	.71
		Hartley	.63	.50	**.57**				**Hartley**	.72	.62	.68
		Bouguet	.59	.45	**.53**				**Bouguet**	.78	.68	**.74**

The only feature extraction technique where distortion corrected images consistently lead to better results is the "fractal dimension: boxcounting" technique. Also, for edge co-occurrence features we notice a few improvements (e.g. for knn classification and Bouguet distortion correction applied to Pars Descendens images also for the other two classifiers). For the other four feature extraction techniques improvements are sParse and in the majority of configurations result degradations are observed.

There are absolutely no trends which justify the much more complicated parameter-free distortion correction as compared to Bouguets software. The tendency that Pars Descendens imagery is more difficult to classify can be confirmed with the results in this study. In most cases SVM classification delivers the best results, but there are also

Table 3. Classification performance for grayscale co-occurrence and SSD feature extraction

Gray scale Co-occurence features							SSD features					
class.	region	images	Sens.	Spec.	Acc.		class.	region	images	Sens.	Spec.	Acc.
diag quad.	Bulbus	distorted	.70	.79	.75		diag. quad.	Bulbus	distorted	.53	.78	.67
		Hartley	.75	.76	**.76**				Hartley	.49	.79	.66
		Bouguet	.77	.81	**.79**				Bouguet	.26	.91	.63
	Pars Desc	distorted	.74	.59	.68			Pars Desc	distorted	.57	.67	.62
		Hartley	.61	.72	.66				Hartley	.71	.54	**.63**
		Bouguet	.67	.67	.67				Bouguet	.57	.71	**.63**
knn	Bulbus	distorted	.75	.79	.77		knn	Bulbus	distorted	.82	.83	.83
		Hartley	.62	.83	.74				Hartley	.73	.82	.78
		Bouguet	.64	.77	.71				Bouguet	.75	.83	.79
	Pars Desc	distorted	.70	.68	.69			Pars Desc	distorted	.71	.67	.69
		Hartley	.68	.65	.67				Hartley	.71	.66	.69
		Bouguet	.68	.54	.62				Bouguet	.65	.69	.67

Table 4. Classification performance for edge co-occurrence and edge orientation features

Edge Co-occurrence features							Edge orientation features					
class.	region	images	Sens.	Spec.	Acc.		class.	region	images	Sens.	Spec.	Acc.
diag quad.	Bulbus	distorted	.51	.67	.60		diag. quad.	Bulbus	distorted	.69	.68	.68
		Hartley	.59	.58	.58				Hartley	.68	.72	**.70**
		Bouguet	.51	.58	.55				Bouguet	.73	.72	**.72**
	Pars Desc	distorted	.45	.67	.55			Pars Desc	distorted	.49	.57	.53
		Hartley	.49	.59	.52				Hartley	.44	.55	.49
		Bouguet	.53	.65	**.58**				Bouguet	.37	.57	.46
knn	Bulbus	distorted	.39	.57	.49		knn	Bulbus	distorted	.85	.45	.62
		Hartley	.48	.55	**.52**				Hartley	.72	.39	.53
		Bouguet	.45	.54	**.50**				Bouguet	.81	.45	.57
	Pars Desc	distorted	.53	.38	.46			Pars Desc	distorted	.45	.52	.51
		Hartley	.54	.48	**.51**				Hartley	.48	.57	**.52**
		Bouguet	.56	.48	**.53**				Bouguet	.58	.43	.51
SVM linear	Bulbus	distorted	.68	.78	.74		SVM quadratic	Bulbus	distorted	.77	.62	.69
		Hartley	.56	.72	.56				Hartley	.63	.56	.59
		Bouguet	.66	.76	.72				Bouguet	.63	.58	.60
	Pars Desc	distorted	.64	.45	.55			Pars Desc	distorted	.59	.55	.57
		Hartley	.61	.40	.52				Hartley	.46	.42	.44
		Bouguet	.63	.49	**.57**				Bouguet	.62	.56	**.60**

a few exceptions. The best result with 0.83 accuracy is obtained with locally invariant fractal features and SSD on the distorted Bulbus image set using SVM classification and knn classification, respectively. Overall, classification accuracy is found to be lower for the considered set of feature descriptors as compared to transform-based or LBP-related methods.

4   Conclusion

Distortion correction does not improve classification of celiac-disease related duodenal image material in many cases, even if edge- and shape-related feature descriptors are used. The role of interpolation as used in all distortion correction techniques needs to be investigated in more detail – especially in the corner regions of the images, where distortion correction is most crucial due to the strong barrel distortion, interpolation

artefacts are most severe due to the large extent of distances to be corrected. Furthermore, the validity of the conclusions found so far needs to be checked for other types of endoscopes (e.g. high-magnification or high-definition endoscopes) and other types of classification tasks (e.g. colon polyp classification, stomach mucosa classification etc.).

References

1. Liedlgruber, M., Uhl, A.: Computer-aided decision support systems for endoscopy in the gastrointestinal tract: A review. IEEE Reviews in Biomedical Engineering 4, 73–88 (2012)
2. Borcharrt, T., Conci, A., d'Ornellas, M.: A warping based approach to correct distortions in endoscopic images. In: Proceedings of the 22nd Brazilian Symposium on Computer Graphics and Image Processing (Sibgrapi 2009), Rio de Janeiro, Brazil (October 2009)
3. Asari, K.V., Kumar, S., Radhakrishnan, D.: A new approach for nonlinear distortion correction in endoscopic images based on least squares estimation. IEEE Transactions on Medical Imaging 18(4), 345–354 (1999)
4. Liedlgruber, M., Uhl, A., Vécsei, A.: Statistical analysis of the impact of distortion (correction) on an automated classification of celiac disease. In: Proceedings of the 17th International Conference on Digital Signal Processing (DSP 2011), Corfu, Greece (July 2011)
5. Gschwandtner, M., Liedlgruber, M., Uhl, A., Vécsei, A.: Experimental study on the impact of endoscope distortion correction on computer-assisted celiac disease diagnosis. In: Proceedings of the 10th International Conference on Information Technology and Applications in Biomedicine (ITAB 2010), Corfu, Greece (November 2010)
6. Hartley, R., Kang, S.: Parameter-free radial distortion correction with center of distortion estimation. IEEE Transactions on Pattern Analysis and Machine Intelligence 29(8), 1309–1321 (2007)
7. Vécsei, A., Amann, G., Hegenbart, S., Liedlgruber, M., Uhl, A.: Automated marsh-like classification of celiac disease in children using an optimized local texture operator. Computers in Biology and Medicine 41(6), 313–325 (2011)
8. Ciaccio, E.J., Tennyson, C.A., Lewis, S.K., Krishnareddy, S., Bhagat, G., Green, P.H.: Distinguishing patients with celiac disease by quantitative analysis of videocapsule endoscopy images. Computer Methods and Programs in Biomedicine 100(1), 39–48 (2010)
9. Barreto, J., Swaminathan, R., Roquette, J.: Non parametric distortion correction in endoscopic medical images. In: Proceedings of the 3DTV Conference 2007, Kos, Greece, pp. 1–4 (2007)
10. Backes, A.R., Bruno, O.M.: A New Approach to Estimate Fractal Dimension of Texture Images. In: Elmoataz, A., Lezoray, O., Nouboud, F., Mammass, D. (eds.) ICISP 2008. LNCS, vol. 5099, pp. 136–143. Springer, Heidelberg (2008)
11. Varma, M., Garg, R.: Locally invariant fractal features for statistical texture classification. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV 2007), Rio de Janeiro, Brazil, pp. 1–8 (2007)
12. Haralick, R., Shanmugan, K., Dinstein, I.: Textual features for image classification. IEEE Transactions on Systems, Man, and Cybernetics SMC-3, 610–621 (1973)
13. Rautkorpi, R., Iivarinen, J.: A Novel Shape Feature for Image Classification and Retrieval. In: Campilho, A., Kamel, M. (eds.) ICIAR 2004. LNCS, vol. 3211, pp. 753–760. Springer, Heidelberg (2004)
14. Manjunath, B., Salembier, P., Sikora, T.: Introduction to MPEG-7: Multimedia Content Description Interface. Wiley & Sons (2002)
15. Ayala, G., Domingo, J.: Spatial size distributions: Applications to shape and texture analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 23(12), 1430–1442 (2001)

Gaussian Process Inference for Estimating Pharmacokinetic Parameters of Dynamic Contrast-Enhanced MR Images

Shijun Wang[1], Peter Liu[1], Baris Turkbey[2], Peter Choyke[2], Peter Pinto[2], and Ronald M. Summers[1]

[1] Imaging Biomarkers and Computer-Aided Diagnosis Laboratory, Radiology and Imaging Sciences, Clinical Center, National Institutes of Health, Bldg 10, Room 1C224, Bethesda, MD 20892-1182, U.S.
{wangshi,liupr}@cc.nih.gov, rms@nih.gov
[2] Molecular Imaging Program, National Cancer Institute, National Institutes of Health, Bldg 10, Room B3B69F, Bethesda, MD 20892-1088, U.S.
{Ismail_Turkbey,pchoyke,peter_pinto}@nih.gov

Abstract. In this paper, we propose a new pharmacokinetic model for parameter estimation of dynamic contrast-enhanced (DCE) MRI by using Gaussian process inference. Our model is based on the Tofts dual-compartment model for the description of tracer kinetics and the observed time series from DCE-MRI is treated as a Gaussian stochastic process. The parameter estimation is done through a maximum likelihood approach and we propose a variant of the coordinate descent method to solve this likelihood maximization problem. The new model was shown to outperform a baseline method on simulated data. Parametric maps generated on prostate DCE data with the new model also provided better enhancement of tumors, lower intensity on false positives, and better boundary delineation when compared with the baseline method. New statistical parameter maps from the process model were also found to be informative, particularly when paired with the PK parameter maps.

Keywords: DCE-MRI, Gaussian Stochastic Process, Pharmacokinetic Model, Bayesian Inference, Coordinate Descent Optimization.

1   Introduction

Dynamic contrast-enhanced MR imaging (DCE-MRI) is a special magnetic resonance imaging (MRI) technique which assesses the micro-vascular status of tissue by repeated acquisition of rapid T_1-weighted images on a region-of-interest (ROI) before, during and after the injection of a low molecular weight contrast agent [1, 2]. The tracking of contrast agent, typically a gadolinium (Gd) compound, provides a way to analyze the pharmacokinetics of the contrast agent which reveals information about the local vascular permeability, blood flow and extracellular volumes in the ROI.

The most widely used pharmacokinetic (PK) model in DCE is the Tofts' dual-compartment model [3], which models the exchange of contrast agent between the vascular space (blood plasma) and the extravascular-extracellular space (EES). The

Tofts' model predicts for two quantitative parameters, K_{trans} and k_{ep}, which have been shown to correlate with cancer (K_{trans} and k_{ep} vary according to the organ being studied). These parameters are typically calculated by using nonlinear regression curve-fitting techniques [4], which fit the concentration profiles of each MRI voxel independently, on a voxel-by-voxel basis. To challenge this voxel-independence assumption, Schmid et al. utilized adaptive Gaussian Markov random fields to estimate the kinetic parameters of DCE-MRI. In this analysis, neighboring voxels in the kinetic parameter estimation reduced the variability in local tumor regions and keep the boundaries between heterogeneous tissues sharp. This neighboring voxel approach was further developed with a spatial prior model by Kelm et al. [5].

Other probability models have also been used to treat the PK modeling problem. For example, Orton et al. utilize a full Bayesian approach [6]. Also, in the work of Chen et al., pixel-wise partial volume effect (PVE) was taken into account due to limited spatial resolution of DCE-MRI and tumor tissue heterogeneity [7]. However, in all of the probability models mentioned, it is assumed that the DCE-MRI observations at different time points are independent. This assumption does not hold for the real physical pharmacokinetic process because the observations from different time points have correlations imposed by the tracer dynamics across time. In reality, time dependence in DCE-MRI exists, and is difficult to characterize. In this paper, to capture the covariance between different time points of DCE-MRI, we treat the time series of DCE-MRI as a stochastic process and use a Gaussian process to describe it [8]. With the Gaussian process modeling, we estimated the pharmacokinetic and statistical parameters using maximum likelihood Bayesian inference.

The paper is organized as following: in Sec. 2, we first introduce the Tofts dual-compartment model briefly, and then propose a new method for the pharmacokinetic parameter estimation using Gaussian process inference (GPI); in Sec. 3 we show experimental results on both simulated data and the prostate MRI dataset; we conclude our findings in this study in Sec. 4 with a short discussion and future direction.

2   Methods

2.1   Pharmacokinetic Model

In this paper, we adopted a dual-compartment model proposed by Tofts et al. which was widely used to describe the dynamic uptake of contrast agent Gd-DTPA into the extracellular-extravascular space (EES) [3] (our method is a general framework and can also utilize other PK models). The concentration of Gd-DTPA as time goes by is modeled by the following equation:

$$C_t(x,t) = K^{trans}\left[C_p(x,t) \otimes \exp(-k_{ep}t) \right] \tag{1}$$

where $C_t(x,t)$ is the measured Gd-DTPA concentration in the tissue located at x at time t, $C_p(x,t)$ is an arterial input function (AIF) giving the tracer concentration in blood plasma, K_{trans} is the transfer constant between the blood plasma and EES, and K_{ep} is the rate constant between the EES and blood plasma. The units of EES and blood plasma are both $\min^{-1}$. $\otimes$ denotes the convolution operator.

By using T_1 weighted images, the Gd-DTPA concentration in tissue $C_t(x,t)$ can be calculated from the following equation [9]:

$$C_t(x,t) = r_1^{-1}\left[T_1(x,t)^{-1} - T_{10}(x)^{-1}\right]$$

(2)

where T_{10} is the T_1 value before contrast agent administration, and $r_1 = 4.24 \ l/s/mmol$ [3] is the longitudinal relaxivity of Gd-DTPA.

The Gd-DTPA concentration $C_p(x,t)$ in blood plasma is modeled by a bi-exponential function (arterial input function (AIF)) [3]:

$$C_p(x,t) = D\sum_{i=1}^{2} a_i \exp(m_i t)$$

(3)

where $a_1 = 3.99 \ kg/l$, $a_1 = 4.78 \ kg/l$, $m_1 = 0.144 \ \min^{-1}$, $m_2 = 0.0111 \ \min^{-1}$ [10], and D is the actual dose of contrast agent for a given body weight ($D = 0.1 \ mmol/kg$ in study [4]). In eq. (3) the first term models the short-term exchange of contrast agent with the tissues; the second term models the removal of contrast agent by the kidneys.

After substituting $C_p(x,t)$ in eq. (1) with eq. (3) and convoluting, we get:

$$C_t(x,t) = DK^{trans}\left(a_1 \frac{\exp(-k_{ep}t) - \exp(-m_1 t)}{m_1 - k_{ep}} + a_2 \frac{\exp(-k_{ep}t) - \exp(-m_2 t)}{m_2 - k_{ep}}\right)$$

(4)

2.2   Gaussian Process Inference

Due to measurement noise in MR imaging and other sources of noise, the observation of tracer concentration in tissue on location x at different time points $t = 1, 2, ..., T$ which is defined as $\mathbf{O}_t(x) = \left[O_t(x,1), O_t(x,2), ..., O_t(x,T)\right]$ is a sequence of random variables, i.e., $O_t(x,t_i) = C_t(x,t_i) + N(0,\sigma_n^2)$. We assume $\mathbf{O}_t(x)$ is a Gaussian stochastic process [8]:

$$\mathbf{O}_t(x) \sim GP\left(\mathbf{C}_t(x), \mathbf{\Sigma}(x)\right),$$

(5)

where $\mathbf{C}_t(x) = \left[C_t(x,1), C_t(x,2), ..., C_t(x,T)\right]$ is the mean function and $\mathbf{\Sigma}(x)$ is the covariance function, $\Sigma_{i,j}(x,t_i,t_j) = \text{cov}\left(C_t(x,t_i), C_t(x,t_j)\right)$.

Based on the observation that for DCE images, the images taken at time points which are close to each other on the time axis should have higher correlation than those taken at time points which are far away from each other on the time axis, we use the "squared exponential" covariance function [8], which fits the data well:

$$\Sigma_{i,j}(x,t_i,t_j) = \sigma_f^2(x)\exp\left[-(t_j - t_i)^2 \Big/ \left(2l(x)^2\right)\right]$$

(6)

where σ_f^2 defines the maximum allowable covariance and l defines the characteristic length-scale (the time required for a $[1-e^{-0.5}] = 39.4\%$ decay in covariation).

By combining the covariance function shown in eq. (6) and the measurement noise with distribution $N\left(0,\sigma_n^2\right)$ from MR imaging together, we use the following covariance function:

$$\Sigma_{i,j}\left(x,t_i,t_j\right) = \sigma_f^2\left(x\right)\exp\left[-\left(t_j - t_i\right)^2 \Big/\left(2l\left(x\right)^2\right)\right] + \sigma_n^2\left(x\right)\delta\left(t_i,t_j\right) \tag{7}$$

where $\delta\left(t_i,t_j\right)$ is the Kronecker delta function. Zero mean Gaussian noise is used for simplicity, though non-zero mean Gaussian noise is easily implementable.

With the Gaussian process assumption, the log likelihood function is:

$$\begin{aligned}
L(x) &= \log p\left(\mathbf{O}_t\left(x\right)\big|\mathbf{C}_t\left(x\right),K^{trans},k_{ep},\sigma_f,\sigma_n,l\right) \\
&= -\frac{1}{2}\log\left|\mathbf{\Sigma}(x)\right| - \frac{1}{2}\left(\mathbf{O}_t\left(x\right)-\mathbf{C}_t\left(x\right)\right)^T\mathbf{\Sigma}(x)^{-1}\left(\mathbf{O}_t\left(x\right)-\mathbf{C}_t\left(x\right)\right) - \frac{T}{2}\log\left(2\pi\right)
\end{aligned} \tag{8}$$

To maximize the likelihood function and find the values of hyperparameters $\left(K^{trans},k_{ep},\sigma_f,\sigma_n,l\right)$ introduced above, we use a variant of coordinate descent [11]; in our variation, the parameters are optimized in two groups: with K_{trans} and K_{ep} in one group, and σ_n, σ_f, and l in the other. Formulas for the gradient follow:

$$\frac{\partial L}{\partial K^{trans}} = \left(\mathbf{O}_t\left(x\right)-\mathbf{C}_t\left(x\right)\right)^T\mathbf{\Sigma}(x)^{-1}D\left[a_1\frac{\exp\left(-k_{ep}t_i\right)-\exp\left(-m_1 t_i\right)}{m_1 - k_{ep}} + a_2\frac{\exp\left(-k_{ep}t_i\right)-\exp\left(-m_2 t_i\right)}{m_2 - k_{ep}}\right]_{i=1,\dots,T}$$

$$\frac{\partial L}{\partial k_{ep}} = \left(\mathbf{O}_t\left(x\right)-\mathbf{C}_t\left(x\right)\right)^T\mathbf{\Sigma}(x)^{-1}DK^{trans}\times$$

$$\left[a_1\frac{\left(\exp\left(-k_{ep}t_i\right)-\exp\left(-m_1 t_i\right)\right)-t_i\exp\left(-k_{ep}t_i\right)\left(m_1-k_{ep}\right)}{\left(m_1-k_{ep}\right)^2} + a_2\frac{\left(\exp\left(-k_{ep}t_i\right)-\exp\left(-m_2 t_i\right)\right)-t_i\exp\left(-k_{ep}t_i\right)\left(m_2-k_{ep}\right)}{\left(m_2-k_{ep}\right)^2}\right]_{i=1,\dots,T}$$

$$\frac{\partial L}{\partial \sigma_f^2\left(x\right)} = -\frac{1}{2}Tr\left(\mathbf{\Sigma}(x)^{-T}\frac{\partial\mathbf{\Sigma}(x)}{\partial\sigma_f^2\left(x\right)}\right) + \frac{1}{2}\left(\mathbf{O}_t\left(x\right)-\mathbf{C}_t\left(x\right)\right)^T\times\mathbf{\Sigma}(x)^{-1}\frac{\partial\mathbf{\Sigma}(x)}{\partial\sigma_f^2\left(x\right)}\mathbf{\Sigma}(x)^{-1}\left(\mathbf{O}_t\left(x\right)-\mathbf{C}_t\left(x\right)\right)$$

$$\frac{\partial\mathbf{\Sigma}_{i,j}\left(x,t_i,t_j\right)}{\partial\sigma_f^2\left(x\right)} = \exp\left[\frac{-\left(t_j-t_i\right)^2}{2l\left(x\right)^2}\right]$$

$$\frac{\partial L}{\partial l\left(x\right)} = -\frac{1}{2}Tr\left(\mathbf{\Sigma}(x)^{-T}\frac{\partial\mathbf{\Sigma}(x)}{\partial l\left(x\right)}\right) + \frac{1}{2}\left(\mathbf{O}_t\left(x\right)-\mathbf{C}_t\left(x\right)\right)^T\times\mathbf{\Sigma}(x)^{-1}\frac{\partial\mathbf{\Sigma}(x)}{\partial l\left(x\right)}\mathbf{\Sigma}(x)^{-1}\left(\mathbf{O}_t\left(x\right)-\mathbf{C}_t\left(x\right)\right)$$

$$\frac{\partial\mathbf{\Sigma}_{i,j}\left(x,t_i,t_j\right)}{\partial l\left(x\right)} = -2\sigma_f^2\left(x\right)\exp\left[\frac{-\left(t_j-t_i\right)^2}{2l\left(x\right)^2}\right]\times\left(-\frac{\left(t_j-t_i\right)^2}{2}\right)\times l\left(x\right)^{-3}$$

$$\frac{\partial L}{\partial \sigma_n^2\left(x\right)} = -\frac{1}{2}Tr\left(\mathbf{\Sigma}(x)^{-T}\frac{\partial\mathbf{\Sigma}(x)}{\partial\sigma_n^2\left(x\right)}\right) + \frac{1}{2}\left(\mathbf{O}_t\left(x\right)-\mathbf{C}_t\left(x\right)\right)^T\times\mathbf{\Sigma}(x)^{-1}\frac{\partial\mathbf{\Sigma}(x)}{\partial\sigma_n^2\left(x\right)}\mathbf{\Sigma}(x)^{-1}\left(\mathbf{O}_t\left(x\right)-\mathbf{C}_t\left(x\right)\right)$$

$$\frac{\partial\mathbf{\Sigma}_{i,j}\left(x,t_i,t_j\right)}{\partial\sigma_n^2\left(x\right)} = \delta\left(t_i,t_j\right) \tag{9}$$

3   Experimental Results and Discussion

The proposed Gaussian process inference (GPI) strategy was studied in two parts. First, the algorithm was tested on simulated data to assess its performance in estimating the "true" local pharmacokinetic parameters. Then, the algorithm was used to generate parametric maps using actual MRI-DCE axial scans of the prostate, and the resulting maps were qualitatively compared with histopathology results from radical prostatectomy after the DCE images were obtained. In both parts of the study, the GPI algorithm was compared against a standard least-squared-error (LSE) minimization method using a Nelder-Mead simplex algorithm. Squared-error-minimization is widely used in DCE-MRI [4].

3.1   Results on Simulated Data

For the simulated DCE data, we generated concentration profiles consisting of 46 Gd-DTPA time-points spaced 5.6 seconds apart (spanning 4.3 minutes). The concentrations were assumed to follow the Gaussian probability distribution described in Sec. 2.2 with covariance terms σ_n equaling 2% of the maximum Gd-DTPA concentration, σ_f equaling 10% of the concentration peak, and l equaling 30 seconds. Simulated data were generated for a physiologically relevant range of 0.1 to 1 min^{-1} for K_{trans} and 0.2 to 3 min^{-1} for K_{ep}. Further, the constraint: $2*K_{trans} < K_{ep} < 10*K_{trans}$ was imposed, allowing for v_e (the EES fraction) to vary between 0.1 and 0.5. For each K_{trans}/K_{ep} pair, 100 independent random samples were generated.
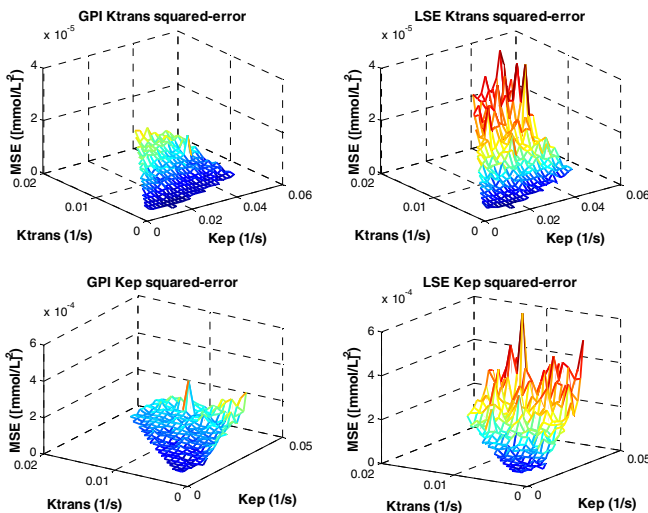
Fig. 1. Simulation results comparing the Mean Squared Errors from the GPI method (left) and the LSE method (right) in K_{trans} (top) and K_{ep} (bottom). On average, the GPI method resulted in lower errors than the LSE method ($p < 0.0001$ for both parameters).

For each sample, the simulated DCE data was fit with both the GPI algorithm and the LSE algorithm. For each algorithm, the absolute error was calculated and averaged over all 100 samples (Figure 1). The results indicate that the GPI algorithm yields significantly lower squared errors than the LSE algorithm for multivariate Gaussian data with non-zero covariance (paired student t-test $p < 0.0001$ for both K_{trans} and K_{ep}).

3.2   Results on Prostate MRI Data

To test the clinical relevance of the GPI method, K_{trans} and K_{ep} maps were generated and compared using GPI and LSE methods. For this study, MR imaging studies of 26 patients were performed using a combination of an endorectal and cardiac coil on a 3T magnet (*Philips Medical Systems*). DCE-MRI and triplanar T2 weighted turbo-spin-echo
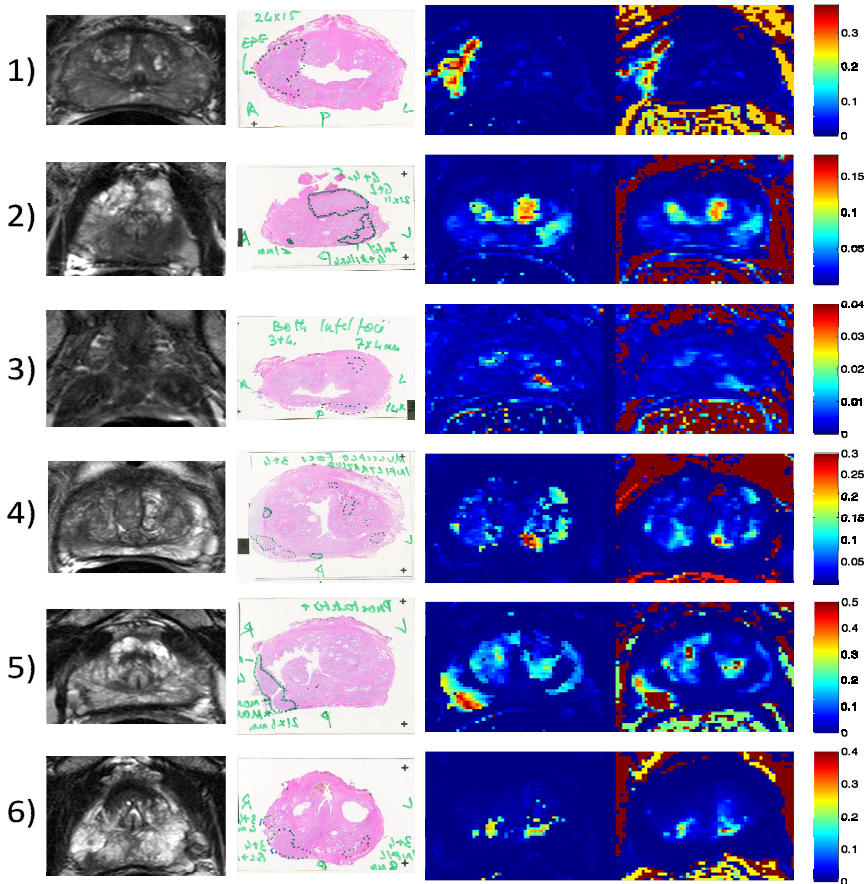
Fig. 2. GPI K_{trans} maps (column 3) are compared with LSE K_{trans} maps (column 4) for six patients. In column 2, the corresponding histopathology labels for prostate cancer (inked in green) are shown. To better show the anatomical boundaries, registered T2-weighted MRI images are also shown (column 1).

data sets were obtained. Following MRI, all patients underwent robotic radical prostatect-omy. The prostate specimens were then sectioned within a customized mold system and mounted on a histopathology slide. Then, the tumors were outlined with their Gleason scores and dimensions in each prostatectomy specimen by two experienced genitourinary pathologists blinded to MRI data.

In Figure 2, parametric maps generated by GPI and LSE are shown next to ground truth histopathology cancer labels. Voxels where convergence was not met were ex-cluded from the analysis. In the cases shown, the GPI maps show promise in increas-ing cancerous lesion enhancement. Particularly, in patients 1, 2, 3, 4 the cancerous regions indicated in the histopathology are more enhanced by GPI than LSE. Further, the GPI maps often reduce the enhancement of false positives shown to be benign. For example, in patient 5, the false-positives in the center of the prostate are darker in the GPI map than in the LSE map. In addition to these findings, it was also deter-mined that the GPI maps tended to yield better outlines of the cancerous lesions. For example, patients 1, 2, 4, 5, and 6 all show a lesion shape in the GPI map that better coincides with the shape found in the histopathology data.

Another major advantage of our model over traditional models is that we can pro-vide additional statistical parameter maps σ_n, σ_f, and l (Fig. 3). These maps reveal more information about tracer exchange kinetics in the tissue and provide comple-mentary information to K_{trans} and K_{ep} maps. Particularly, it was found that the edges of prostate tumor often showed higher σ_f and l values. This coincides with the finding that GPI provides better edge delineation of cancer lesions. For example, patient 1 in Fig. 2 showed a truer contour on the left and right edges of the lesion with GPI than LSE; this corresponds with the higher σ_f, and l values exhibited in the same area in Fig. 3 (i.e. in areas where the covariance terms are high, GPI is expected to outper-form LSE, as demonstrated in Fig. 1). The higher covariance at lesion boundaries is suspected to be caused by tissue heterogeneity.
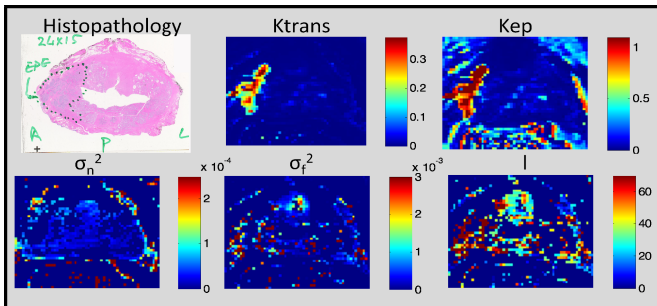
Fig. 3. All parametric maps generated by the GPI algorithm are shown with corresponding histopatholo-gy labels (cancer delineated in green). The color maps used for each parameter are show in the colorbar to the right of the image, with units in seconds^{-1} (K_{trans} and K_{ep}), mmol/L (σ_n^2 and σ_f^2), and seconds (l).

4   Conclusion

In this paper, we propose a new pharmacokinetic model for parameter estimation of DCE-MRI by using Gaussian process inference. The parameter estimation is done

through maximum likelihood approach and we propose a coordinate descent variant to solve this likelihood maximization problem.

By testing the algorithm on simulated MRI data, it was shown that the Gaussian process inference approach results in significantly more accurate results for MRI data that exhibits Gaussian covariance. Further, the K_{trans} maps generated by the GPI algorithm show promise in yielding higher sensitivity and specificities, while leading to better delineations of lesion boundaries. σ_f, and l also showed promise in aiding clinicians to detect tumor boundaries. In the future, quantitative comparisons of GPI with LSE and other PK methods are required to determine the true benefit of GPI.

References

[1] Padhani, A.R., Leach, M.O.: Antivascular cancer treatments: functional assessments by dynamic contrast-enhanced magnetic resonance imaging. Abdominal Imaging 30, 324–341 (2005)

[2] O'Connor, J.P.B., Jackson, A., Parker, G.J.M., Jayson, G.C.: DCE-MRI biomarkers in the clinical evaluation of antiangiogenic and vascular disrupting agents. British Journal of Cancer 96, 189–195 (2007)

[3] Tofts, P.S.: Modeling tracer kinetics in dynamic Gd-DTPA MR imaging. JMRI-Journal of Magnetic Resonance Imaging 7, 91–101 (1997)

[4] Schmid, V.J., Whitcher, B., Padhani, A.R., Taylor, N.J., Yang, G.Z.: Bayesian methods for pharmacokinetic models in dynamic contrast-enhanced magnetic resonance imaging. IEEE Transactions on Medical Imaging 25, 1627–1636 (2006)

[5] Kelm, B.M., Menze, B.H., Nix, O., Zechmann, C.M., Hamprecht, F.A.: Estimating Kinetic Parameter Maps From Dynamic Contrast-Enhanced MRI Using Spatial Prior Knowledge. IEEE Transactions on Medical Imaging 28, 1534–1547 (2009)

[6] Orton, M.R., Collins, D.J., Walker-Samuel, S., d'Arcy, J.A., Hawkes, D.J., Atkinson, D., Leach, M.O.: Bayesian estimation of pharmacokinetic parameters for DCE-MRI with a robust treatment of enhancement onset time. Physics in Medicine and Biology 52, 2393–2408 (2007)

[7] Chen, L., Choyke, P.L., Chan, T.H., Chi, C.Y., Wang, G., Wang, Y.: Tissue-Specific Compartmental Analysis for Dynamic Contrast-Enhanced MR Imaging of Complex Tumors. IEEE Transactions on Medical Imaging 30, 2044–2058 (2011)

[8] Rasmussen, C., Williams, C.: Gaussian Processes for Machine Learning. MIT Press (2006)

[9] Buckley, D., Parker, G.: Measuring contrast agent concentration in T1-weighted dynamic contrast-enhanced MRI. Presented at Dynamic Contrast-Enhanced Magnetic Resonance Imaging in Oncology (2005)

[10] Rohrer, M., Bauer, H., Mintorovitch, J., Requardt, M., Weinmann, H.J.: Comparison of magnetic properties of MRI contrast media solutions at different magnetic field strengths. Investigative Radiology 40, 715–724 (2005)

[11] Bouman, C.A., Sauer, K.: A unified approach to statistical tomography using coordinate descent optimization. IEEE Transactions on Image Processing 5, 480–492 (1996)

Automatic Localization and Identification of Vertebrae in Arbitrary Field-of-View CT Scans

Ben Glocker[1], J. Feulner[2], Antonio Criminisi[1], D.R. Haynor[3],
and E. Konukoglu[1]

[1] Microsoft Research, Cambridge, UK
[2] University of Erlangen-Nuremberg, Erlangen, Germany
[3] University of Washington, Seattle, WA, USA

Abstract. This paper presents a new method for automatic localization and identification of vertebrae in arbitrary field-of-view CT scans. No assumptions are made about which section of the spine is visible or to which extent. Thus, our approach is more general than previous work while being computationally efficient. Our algorithm is based on regression forests and probabilistic graphical models. The discriminative, regression part aims at roughly detecting the visible part of the spine. Accurate localization and identification of individual vertebrae is achieved through a generative model capturing spinal shape and appearance. The system is evaluated quantitatively on 200 CT scans, the largest dataset reported for this purpose. We obtain an overall median localization error of less than 6mm, with an identification rate of 81%.

1 Introduction

This paper proposes an algorithm for automatic detection, localization, and identification of individual vertebrae in computed tomography scans. A variety of tasks beyond spine specific analysis can immediately benefit from such a system. The spine provides a natural patient-specific coordinate system, where individual vertebrae serve as anatomical landmarks. These can be used, for instance, for semantically guided inspection tools, linking of radiological reports with corresponding image regions, or for robust initialization of image registration. Vertebrae localization also provides valuable priors for subsequent tasks such as anatomy segmentation, image retrieval, shape and population analysis.

The challenges associated with automatic localization of individual vertebrae arise from i) the repetitive nature of these structures, ii) the variability of normal and pathological anatomy, iii) and the variability of images (*e.g.* resolution and field-of-view). A common approach for vertebra (in CT) and intervertebral disks (in MRI) is to employ a multi-stage approach. In the first stage a detector in the form of a filter [1, 2], a single/multi-class classifier [3–8] or a model-based Hough transform [9] is used to detect potential vertebra candidates. As these candidates may contain many false positive responses a second stage is applied to add robustness. Prior knowledge on the global shape and/or appearance of individual vertebrae and their interconnections is used. In [2], a clever search is

performed based on prior information through the candidates, while [1, 4, 5] fit a low order polynomial curve to the candidates to remove outliers. In [3, 6, 8, 9] the authors add prior information via graphical models, such as Hidden Markov Models (HMMs) [10], and infer the maximum a-posteriori (MAP) estimate for the vertebrae locations. In contrast, [11, 12] use a fully generative model, and inference is achieved via generalized expectation-maximization, while in [7] deformable templates are used for segmentation and subsequent identification.

Although previous works achieve high localization accuracy they cannot handle completely general scans where it is not known in advance which portion of the anatomy is visible. In fact, most algorithms require a priori knowledge of which vertebrae are visible in the scan [1–8]. They either focus on a specific region, such as lumbar or thoracic, or need to modify their models based on the expected spine region. In [11, 12] approximate alignment between scans is assumed. To the best of our knowledge the only work that explicitly handles arbitrary scans is [9]. However, the added generality comes at high computational cost. Based on an affine vertbra registration algorithm, the identification phase is reported to take up to 36 minutes for 12 thoracic vertebrae.

In this paper we overcome those drawbacks with a vertebra localization and identification algorithm which is both robust and efficient. Its main advantage is the automatic handling of arbitrary field-of-view scans displaying widely varying anatomical regions. For instance, in a narrow abdominal scan we may be able to see just a handful of vertebrae together with the kidneys. A radiologist makes use of such contextual information to infer that we are looking at a lumbar section of the spine. In our system we incorporate contextual information within a regression forest algorithm. More specifically, we build upon state-of-the-art supervised, non-linear regression techniques [13] used jointly with a probabilistic, generative prior of spinal shape and appearance. The forest provides context-aware, fast estimation of vertebrae centres. In a second stage a joint model of vertebra appearance and global spine shape yields a refined localization as well as individual vertebra identification. The whole process takes less than 2 minutes on a standard desktop machine, thus allowing integration in existing image analysis pipelines. Details of our approach are presented in the next section, followed by an extensive quantitative validation on a large labelled dataset of 200 CT scans.

2 Vertebrae Localization and Identification

Similar to previous methods, our system relies on a two-stage approach. The first stage aims at roughly detecting and localizing all vertebrae of the spine within the image. Refinement of vertebrae positions and their identification is obtained in the second stage.

Previous work extracts location candidates via classification (*e.g.* in a sliding window framework). In contrast, here we take a more direct, regression approach. In fact, given a training set we learn a regression function which associates vertebrae positions with image points, directly. By combining the predictions of many (possibly sparse) sampled image points, one can obtain robust and efficient

location estimates while avoiding sliding window-like expensive search. Note that these location estimates are not restricted to be inside the visible image domain, and thus, an approximate localization of *all* vertebrae is possible, even if only a small portion of the patient's anatomy is visible.

Contextual reasoning is enabled via long-range spatial features, like the ones used in [14, 15]. This way, the presence of organs such as kidneys, liver, or lungs provide strong indications about the presence of certain vertebrae. Regression forests enable us to select automatically the most discriminative features of accurate prediction. Next, we first formalize the regression technique. Then we describe the second, refinement stage.

2.1   Stage 1: Regression Forests

Regression forests is a supervised learning technique for the probabilistic estimation of continuous variables. Recent work has shown that this technique can be successfully applied to organ bounding box localization in CT [14] and MR [15]. In our application, we aim at regressing the set of n vertebrae centroids denoted as $\mathcal{C} = \{\mathbf{c}_i\}_1^n$ with $\mathbf{c} \in \mathbb{R}^3$. The predictor function is then defined as $p(\mathcal{C}|\mathcal{X})$ where $\mathcal{X} = \{(\mathbf{x}_j, \mathbf{f}_j)\}$ is a set of pairs of feature vectors $\mathbf{f} = (f_1, \ldots, f_d) \in \mathbb{R}^d$ with visual feature responses f extracted for individual image points $\mathbf{x} \in \mathbb{R}^3$. Thus, given the data $\mathcal{X}$ obtained from an image this discriminative predictor allows to estimate the most likely positions of vertebrae in that image.

Regression forests tackle the problem of learning the predictor in a divide-and-conquer fashion. A forest is an ensemble of T (probabilistic) binary decision trees, where each tree t learns its own predictor $p_t(\mathcal{C}|\mathcal{X})$. Given a training set $\mathcal{T} = \{(\mathcal{X}_k, \mathcal{C}_k)\}$, obtained from annotated CT scans, training a tree is done by successively subdividing the training examples within the feature space. At each internal node data subsets $\mathcal{T}_\mathrm{L}, \mathcal{T}_\mathrm{R}$ are sent to the left and right child node. A local split function is determined at each node based on the arriving examples. The splits are obtained by (randomly) selecting one feature response for all examples and optimizing over a threshold w.r.t. an objective function. The splitting aims at clustering examples in leaf (terminal) nodes with both consistent annotations and similar feature responses. Tree growing stops when a certain tree depth is reached. In order to extract visual feature responses $\mathbf{f}$, we employ displaced box features which: i) capture long-range appearance context [14, 15] and, ii) can be implemented efficiently via integral image processing [16]. Injecting randomness during the tree training process decreases correlation between individual trees and increases the forest generalization capabilities[1].

Forest Training. Since our training set is composed of feature vectors for image points from arbitrary, unregistered CT scans with varying resolutions and croppings, a regression over absolute image coordinates of vertebrae is not meaningful. Instead, and similar to the case of bounding box regression [14, 15], we associate each training point $\mathbf{x}$ with its relative displacements $\{\mathbf{d}_i\}$, *i.e.* the offsets to *all* available vertebrae centroids given by $\mathbf{d}_i = \mathbf{c}_i - \mathbf{x}$. We employ multivariate

---

[1] More details on forests can be found in [13].
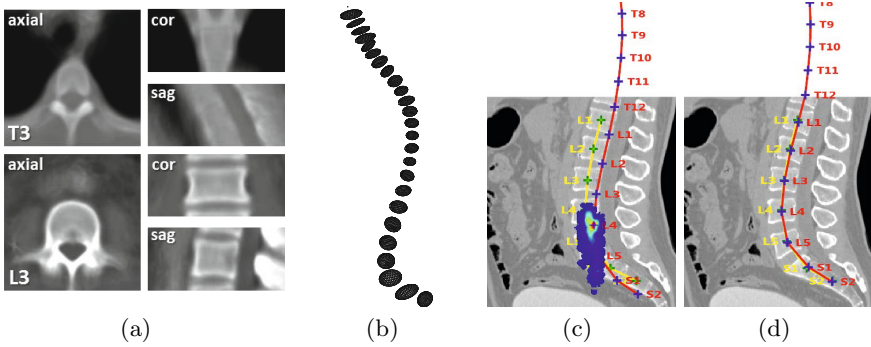
Fig. 1. (a) Mean images of our appearance model for T3 and L3. (b) Visualization of Gaussian densities for offset probabilities in our shape model. Ellipses illustrate one standard deviation w.r.t. covariance matrices. (c) Output of the regression forest for a test image (red) with an overlay of expert annotation (yellow) and prediction distribution for L4. (d) Result after refinement via HMM. Besides accurate predictions for vertebrae within the image, our method yields reasonable predictions outside.

Gaussians to model the node predictor functions: $\mathcal{N}(\mu, \Sigma | \mathcal{D})$ with $\mu \in \mathbb{R}^{3n}$ where $\mathcal{D} = \{\{\mathbf{d}_i\}_j\}$ is the set of offsets obtained from all training points arriving at that node. The training objective function is defined as $\xi(\mathcal{T}_{\mathrm{L}}, \mathcal{T}_{\mathrm{R}}) = \mathrm{tr}(\Sigma_{\mathrm{L}}) + \mathrm{tr}(\Sigma_{\mathrm{R}})$. Node training aims at minimizing ξ which, in turn, minimizes the diagonal entries of the covariance matrices. This produces child subsets $\mathcal{T}_{\mathrm{L,R}}$ with lower uncertainty and higher confidence in the prediction of vertebrae location.

Forest Testing. Given a previously unseen CT scan, image points cast a (probabilistic) vote on the position of all vertebrae. In fact, each point is pushed through all trained trees. Each split node then sends the point to its left or right child depending on its feature vector, recursively until the point reaches a leaf node. The corresponding predictor function (*i.e.* Gaussian in this case) is read out and used for making one prediction for all vertebrae positions relative to the image point location. Aggregating all predictions over all trees and test points yields a distribution over all vertebrae positions (see Fig. 1(c)). For robustness, we approximate the maximum a-posteriori estimate $\hat{\mathcal{C}}$ of this distribution through mean-shift (initialized with the maximum response of a low-resolution histogram over predictions with bin size 4mm). The output vertebrae locations obtained here are then used as input for our refinement step, described next.

2.2 Stage 2: Hidden Markov Model

The second stage of our approach aims at refining the localization of all centroids of vertebrae visible in the image. To this end, we employ a joint prior model of vertebra appearance and spinal shape. The model parameters are optimized using the same data set employed to train the forest.

Vertebra Appearance Model. The appearance model consists of pairs of mean and variance images $\mathcal{A} = \{(M_i, V_i)\}_1^n$, one pair per vertebrae. These pairs

are computed by super-imposing sub-volumes of size $11 \times 11 \times 5$cm cropped from the training data and centered on each annotated vertebra. A few iterations of nonlinear registration are performed to increase the sharpness of the mean images (see examples in Fig. 1(a)). Given a candidate position $\mathbf{c}_i$ we define a likelihood function w.r.t. the appearance model as

$$p(\mathbf{c}_i|\mathcal{A}) = \int_{\Omega_i} \frac{1}{\sqrt{2\pi V_i(\mathbf{x})}} \exp\left(-\frac{(I(\mathbf{c}_i - \mathbf{x}) - M_i(\mathbf{x}))^2}{2V_i(\mathbf{x})}\right) d\mathbf{x} \quad . \tag{1}$$

Spine Shape Prior. The shape model captures conditional probabilities over vertebrae positions. We determine a set of distributions $\mathcal{S} = \{p(\mathbf{c}_i|\mathbf{c}_{i-1}, s)\}_2^n$ where

$$p(\mathbf{c}_i|\mathbf{c}_{i-1}, s) \hat{=} \mathcal{N}\left(\frac{\|\mathbf{c}_i - \mathbf{c}_{i-1}\|}{s}|\mu_i, \Sigma_i\right) \quad \text{with} \quad s = \frac{1}{n-1}\sum_{i=2}^{n}\frac{\|\mathbf{c}_i - \mathbf{c}_{i-1}\|}{\mathbb{E}(\|\mathbf{c}_i - \mathbf{c}_{i-1}\|)} \tag{2}$$

The variable s corresponds to a global scale factor which reflects overall body size. A visualization of these offset distributions is shown in Fig. 1(b).

Joint Shape and Appearance. We define an HMM with hidden states for each vertebrae position, appearance likelihoods and inter-vertebra shape priors. The HMM distribution $p(\mathcal{C}|\mathcal{A}, \mathcal{S}, s)$ conditioned on global scale yields the energy:

$$E(\mathcal{C}; s) = -\sum_{i=1}^{n}\log\left[p(\mathbf{c}_i|\mathcal{A})\right] - \lambda\sum_{i=2}^{n}\log\left[p(\mathbf{c}_i|\mathbf{c}_{i-1}, s)\right] \quad . \tag{3}$$

Given a value for s and multiple sampled location candidates MAP inference can be achieved via dynamic programming. Several thousand candidate locations are sampled from the vicinity of the forest prediction using a normal distribution $\mathcal{N}(\mathbf{c}_i, \sigma^2)$ with $\sigma = 30$mm. In practice, we optimize over 7 scale parameters from a $[0.85, 1.15]$ interval covering 97% of observed patient scales. The weighting parameter λ controls the influence of the shape term, and thus, how much the solution can deviate from the mean shape. Throughout our experiments this weighting is fixed to 0.1. In the exemplary result in Fig. 1(d) notice how the thoracic vertebrae follow reasonable predictions outside the image domain.

3   Experiments

Our spine model includes $n = 26$ individual vertebrae, where the regular 24 from the cervical, thoracic, and lumbar regions are augmented with 2 centroids denoted as S1 and S2 located on the sacrum. We evaluate accuracy of both localization and identification on a dataset of 200 CT scans where the centroids of all visible vertebrae have been manually selected. The dataset is a heterogeneous collection of CT scans from different clinical centers equipped with varying hardware. Images have been acquired for diverse clinical tasks. The scans vary widely, especially in terms of vertical cropping, image noise and physical resolution. The inter-axial distance varies between 0.5 and 6.5mm, with 79 scans having a distance of 3.75mm. The number of slices varies between 51 and 2058 with an average of about 240. Some highly cropped images show only 4 vertebrae.

Table 1. Summary of the localization and identification errors evaluated on 200 CTs

Vertebrae		Stage 1: Regression Forest			Stage 2: HMM			Distance to Closest			Identification	
Region	Counts	Median	Mean	Std	Median	Mean	Std	Median	Mean	Std	Correct	Rate
All	2595	15.91	18.35	11.32	5.31	9.50	10.55	4.79	6.10	5.53	2089	81%
Cervical	116	25.97	30.74	18.64	6.87	10.85	12.49	6.14	8.53	9.05	84	72%
Thoracic	1417	15.79	18.20	10.81	5.51	9.83	10.44	4.91	5.94	4.84	1100	78%
Lumbar	1062	15.40	17.20	10.07	4.88	8.92	10.45	4.59	6.06	5.82	905	85%
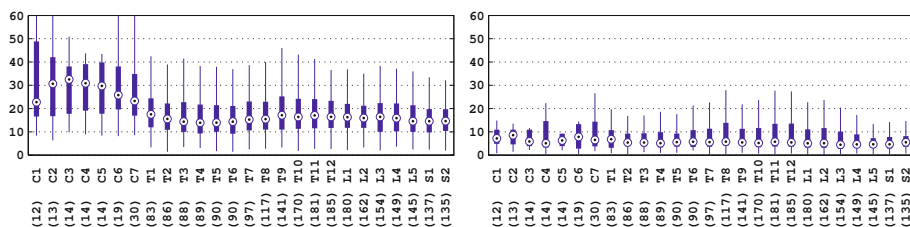
Fig. 2. Error statistics for individual vertebrae: (left) forest prediction only, (right) refinement via HMM. The counts of each vertebra in our database are given in brackets.

3.1 Results

We split the 200 CT scans into two non-overlapping sets with 100 scans each. Each set is used once for both: i) forest training (50 trees, depth 20), and ii) estimating the shape and appearance model; the remaining set is used for testing. Thus we can report errors for all 200 scans and a total of 2595 vertebrae.

Localization Errors defined as distance (in mm) of each predicted vertebra location from its expert annotation are summarized in Tab. 1 and Fig. 2. We obtain a median error of less than 6mm. The highest errors are within the cervical region with a median of about 7mm. This is due to the low number of cervical vertebrae in our data sets (only 6-17 examples for C1 to C7). The lowest errors are obtained for the lumbar region (including S1 and S2) where visual appearance is more discriminative and low image resolution has less of an impact. In Fig. 2, we plot the statistics over localization errors graphically. The figure highlights the massive improvement due to the HMM refinement step. We also give the counts for vertebrae as they appear in the set of 200 scans.

In Tab. 1 we also report distances of predictions from the closest vertebra. So, we have an estimate whether our prediction is in fact located on a vertebra, even if it is not the correct one. The difference between these errors and the ones when considering the correct vertebra is higher for the cervical region, while in the thoracic and lumbar region the difference between median errors is less than 0.6mm. This indicates that in most cases the closest centroid in T and L regions is the correct one, while in C our predicted localizations are in fact on the spine, but might be in some cases on the incorrect vertebra. This confirms that the increased difficulty in discriminating close-by vertebrae in the cervical and upper
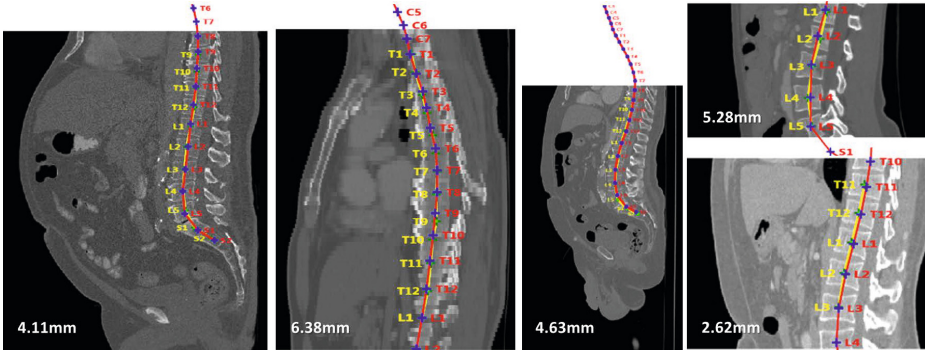
Fig. 3. Our results (red) for varying CT scans (cropped, low resolution, noise, and large field-of-view). Numbers are the mean errors w.r.t. expert annotations (yellow).

thoracic regions contributes to most of our errors. Still, even in these challenging cases our system is able to robustly localize the overall spinal anatomy, which in certain applications might be sufficient.

Identification Errors. We define a vertebra identification criterion as follows: if the closest centroid in the expert annotation corresponds to the correct vertebra, and the localization error is less than 20mm, we call the identification correct. The last two columns in Tab. 1 show an overall success rate of 81%, *i.e.* 2089 out of 2595 vertebrae correctly identified.

Efficiency. In the proposed system training a single tree takes about 3 minutes on randomly sampled 5% of the image points of 100 scans. Each tree can be trained independently and in parallel. More importantly, testing is very fast. In fact, testing a whole forest on a scan takes less than 1 second. The HMM-based refinement takes about 5-15 seconds for each scale s, also depending on the number of vertebrae within the image. Thus, in total, the localization and identification of all vertebrae in one test image is achieved in less than 2 minutes. Figure 3 provides some visual results and corresponding mean errors.

4   Conclusion

This paper has proposed an automatic and efficient approach for the localization and identification of vertebrae in generic CT scans. The algorithm does not make any assumptions on the input images and can deal with highly cropped scans and partially visible spines. Exhaustive experiments on a database of 200 labelled CT scans demonstrate the strength of our joint model of discriminative regression and generative appearance and shape modeling.

In the future, increasing the amount of training data, in particular, for the cervical region would produce an increase in accuracy across the entire spine.

Additionally, automatically predicting the patient overall size could replace the current scale search step and reduce testing times to only a few seconds. Further investigation will also be carried out w.r.t. highly pathological cases of spine such as high-grade scoliosis and cifosis.

References

1. Peng, Z., Zhong, J., Wee, W., Lee, J.H.: Automated Vertebra Detection and Segmentation from the Whole Spine MR Images. In: IEEE EMBC, pp. 2527–2530 (2005)
2. Pekar, V., Bystrov, D., Heese, H.S., Dries, S.P.M., Schmidt, S., Grewer, R., den Harder, C.J., Bergmans, R.C., Simonetti, A.W., van Muiswinkel, A.M.: Automated Planning of Scan Geometries in Spine MRI Scans. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part I. LNCS, vol. 4791, pp. 601–608. Springer, Heidelberg (2007)
3. Schmidt, S., Kappes, J.H., Bergtholdt, M., Pekar, V., Dries, S.P.M., Bystrov, D., Schnörr, 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)
4. Huang, S.H., Lai, S.H., Carol, L.N.: A statistical learning approach to vertebra detection and segmentation from spinal MRI. In: IEEE ISBI, vol. 28, pp. 1595–1605 (2008)
5. Huang, S.H., Chu, Y.H., Lai, S.H., Novak, C.L.: Learning-based Vertebra Detection and Iterative Normalized-Cut Segmentation for Spinal MRI. IEEE TMI 28(10), 1595–1605 (2009)
6. Kelm, B., Zhou, K., Sühling, M., Zheng, Y., Wels, M., Comaniciu, D.: Detection of 3D Spinal Geometry Using Iterated Marginal Space Learning. In: Workshop MedCV (2010)
7. 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)
8. Oktay, A.B., Akgul, Y.S.: Localization of the Lumbar Discs Using Machine Learning and Exact Probabilistic Inference. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part III. LNCS, vol. 6893, pp. 158–165. Springer, Heidelberg (2011)
9. Klinder, T., Ostermann, J., Ehm, M., Franz, A., Kneser, R., Lorenz, C.: Automated Model-based Vertebra Detection, Identification, and Segmentation in CT Images. MedIA 13(3), 471–482 (2009)
10. Bishop, C.: Pattern Recognition and Machine Learning. Springer (2006)
11. Corso, J.J., Alomari, R.S., Chaudhary, V.: Lumbar Disc Localization and Labeling with a Probabilistic Model on Both Pixel and Object Features. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 202–210. Springer, Heidelberg (2008)
12. Alomari, R., Corso, J., Chaudhary, V.: Labeling of Lumbar Discs using both Pixel- and Object-level Features with a Two-Level Probabilistic Model. IEEE TMI 30(1), 1–10 (2011)

13. Criminisi, A., Shotton, J., Konukoglu, E.: Decision Forests: A Unified Framework. Foundations and Trends in Computer Graphics and Vision 7(2-3) (2011)
14. Criminisi, A., Shotton, J., Robertson, D., Konukoglu, E.: Regression Forests for Efficient Anatomy Detection and Localization in CT Studies. In: Workshop MedCV (2010)
15. Pauly, O., Glocker, B., Criminisi, A., Mateus, D., Möller, A.M., Nekolla, S., Navab, N.: Fast Multiple Organ Detection and Localization in Whole-Body MR Dixon Sequences. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part III. LNCS, vol. 6893, pp. 239–247. Springer, Heidelberg (2011)
16. Viola, P., Jones, M.J.: Robust Real-Time Face Detection. IJCV 57(2), 137–154 (2004)

Pathology Hinting as the Combination of Automatic Segmentation with a Statistical Shape Model

Pascal A. Dufour[1,2], Hannan Abdillahi[3], Lala Ceklic[3],
Ute Wolf-Schnurrbusch[2,3], and Jens Kowal[1,2]

[1] ARTORG Center for Biomedical Engineering Research, Ophthalmic Technologies,
University of Bern, 3010 Bern, Switzerland
`pascal.dufour@artorg.unibe.ch`
[2] University Hospital Bern, Ophthalmic Department, 3010 Bern, Switzerland
[3] Bern Photographic Reading Center, 3010 Bern, Switzerland

Abstract. With improvements in acquisition speed and quality, the amount of medical image data to be screened by clinicians is starting to become challenging in the daily clinical practice. To quickly visualize and find abnormalities in medical images, we propose a new method combining segmentation algorithms with statistical shape models. A statistical shape model built from a healthy population will have a close fit in healthy regions. The model will however not fit to morphological abnormalities often present in the areas of pathologies. Using the residual fitting error of the statistical shape model, pathologies can be visualized very quickly. This idea is applied to finding drusen in the retinal pigment epithelium (RPE) of optical coherence tomography (OCT) volumes. A segmentation technique able to accurately segment drusen in patients with age-related macular degeneration (AMD) is applied. The segmentation is then analyzed with a statistical shape model to visualize potentially pathological areas. An extensive evaluation is performed to validate the segmentation algorithm, as well as the quality and sensitivity of the hinting system. Most of the drusen with a height of 85.5μm were detected, and all drusen at least 93.6μm high were detected.

Keywords: pathology hinting, statistical shape model, multi-surface segmentation, optical coherence tomography.

1 Introduction

With the recent advances in OCT, the quality and acquisition speed has increased dramatically. This has revolutionized ophthalmology, as it allows the fast and non-invasive imaging of various structures of the human eye. Today, the imaging of the retinal layers with OCT is standard clinical practice. As with other image modalities in medicine, the faster acquisition speed and increased image resolution also increases the amount of work to be done by the clinician to screen the datasets and state a diagnosis. The current clinical practice of analyzing an OCT volume is visual inspection of each individual B-scan. As such an

inspection takes minutes instead of seconds, in practice clinicians only quickly scroll through the B-scans and abnormalities are often missed. This calls for new analysis and visualization tools.

Segmentation of OCT datasets has become an important tool to quantitatively analyze retinal layers. Recent advances in segmentation techniques [1], [2] make it possible to segment the drusen of AMD patients with high accuracy.

The main contribution of this work is the use of a statistical model as a tool to quickly visualize possibly pathological regions in a dataset. For this we built a statistical shape model [3] from segmentations of fovea-centered OCT datasets of healthy patients. To analyze a new patient's OCT volume, the dataset is first segmented and the statistical shape model is then fitted to that segmentation. The error of the fitting is then visualized as a top down rendering (en-face map).

Key Contributions: We propose the combination of a graph-based segmentation algorithm with a statistical shape model to build a quick pathology hinting system. Our key contributions are twofold:

- A graph-based method for highly accurate drusen segmentation
- A hinting system able to quickly visualize morphological abnormalities.

2   Methods

2.1   Segmentation

The segmentation is based on the optimal net surface problems introduced by Wu et al. [4] and extended to multiple surfaces by Li et al. [5]. Garvin et al. [1] applied this algorithm to the segmentation of retinal layers in OCT volumes. Additional soft constraints, proposed by Song et al. [2], were applied. These constraints allow us to add costs for the rigidity of a surface and costs for the distance between two surfaces. Prior information can therefore be better incorporated into the graph, resulting in an improved segmentation [2].

Furthermore, the soft constraints also enable more possibilities when segmenting pathologies such as drusen. Drusen are an accumulation of extracellular material in the Bruch's membrane (BM) of the retina [6]. The result is a displacement of the cell layers above it. In the OCT, this is most visible in the displacement of the inner and outer photoreceptor segments (IS/OS). However, the boundary between the choroid and the Bruch's membrane is left mostly intact. This makes it easy to segment the lower BM boundary with this method. See Fig. 1(b) for an example of large drusen in an OCT. By adding a strong soft constraint on the rigidity of the surface of the lower BM boundary, it is possible to get a smooth and accurate segmentation.

The next step is the segmentation of the upper IS/OS boundary. Because the displacement from the drusen can be quite large, only a weak soft constraint on the rigidity is added. Additionally, a medium soft constraint on the expected distance to the lower BM boundary is added. This favors a segmentation that is close to the expected IS/OS position of a healthy RPE, but still allows the segmentation of the displaced IS/OS when a druse is present. Fig. 1 shows two examples of the applied segmentation.

(a)   (b)
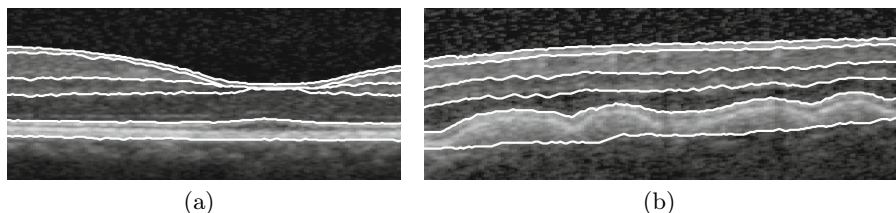
Fig. 1. Segmented OCT B-scans: (a) Healthy retina around the fovea, (b) AMD patient with drusen

2.2 Statistical Shape Model

Statistical shape models can capture the natural shape variation from training shapes [3]. The idea is to allow the statistical shape model to generate the whole range of healthy shapes, without being able to generate pathological shapes. By fitting the statistical shape model to a new segmentation, we will then be able to detect if and where the new segmentation cannot be accurately represented by this model.

Model Building. To build a statistical shape model, the training shapes first have to be brought into a common coordinate frame. Because we do not use 3D positions to build the model, but only the differences between the segmented layers, the rotation of the shapes can be omitted and a translation is enough to align all training shapes. The fovea was used as an anchor, which was detected by finding the lowest point on the segmentation of the inner limiting membrane (ILM). All shapes were translated so the fovea is at the coordinate origin $(0,0)$. Each position $(i \cdot d_x, j \cdot d_y)$ relative to the fovea becomes a landmark position, where i and j are integers with predefined boundaries and d_x and d_y are the sampling spacing in x- and y-direction. This simplifies the landmarking process, as the landmark positions form a simple grid around the fovea and no anatomical landmarks are required. The statistical shape model was then built by computing the mean shape and covariance matrix, and principal component analysis (PCA) was applied for dimensionality reduction [3]. Automatic segmentations of 28 OCT volumes were used as training shapes. Fig. 2 illustrates all steps required to build the statistical shape model.

Model Fitting. Given a new segmentation of a new OCT volume, the statistical shape model can be deformed so that it minimizes the distance to the shape vector of that new segmentation. See [3] for an iterative approach to deform the statistical shape model to a new shape vector.

During the fitting, the deformation of the model is limited so it is able to only represent about 99% of the variation encountered in the training datasets. This ensures that the generated shape is similar to the shapes seen in the training datasets and cannot deform to an extreme shape. See Fig. 3 for the necessary steps to fit the model to a new segmentation.
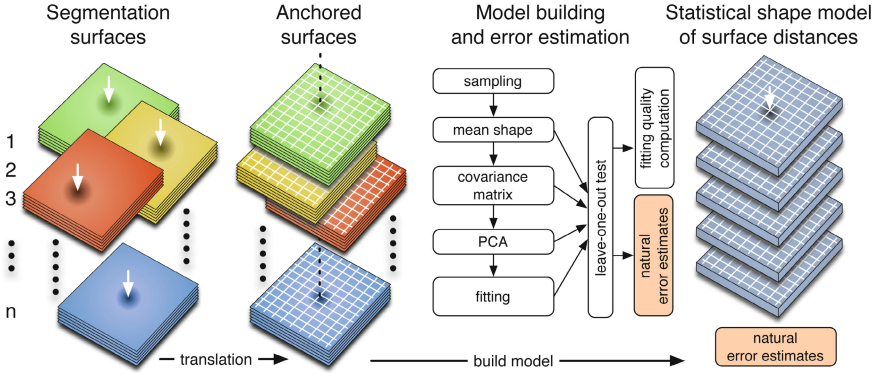
Fig. 2. Complete statistical shape model building procedure including leave-one-out test for the natural error estimate computation

2.3   Pathology Hinting

For every landmark, the residual absolute fitting error between the deformed model and the new segmentation serves as a measure of pathology at that position. The hinting can be further improved by normalizing that error.

When building the statistical shape model, we perform a leave-one-out test to estimate the natural residual errors. For every landmark position, the distribution of the mean unsigned error ε_μ and its variance ε_σ^2 is computed. Assuming a normal distribution of that error, we now know that 68% of the errors in the training set are within the interval $\varepsilon_\mu \pm \varepsilon_\sigma$, 95% are within $\varepsilon_\mu \pm 2\varepsilon_\sigma$, and so on.

Let's say we now fit the model to a new segmentation and encounter a residual fitting error ε at a specific landmark, for example $\varepsilon = \varepsilon_\mu + 3\varepsilon_\sigma$. We know that at that landmark, only 0.27% of the measured errors in a healthy dataset are at least this large and that therefore the landmark is highly abnormal. We formulate the measure of abnormality ψ at a specific landmark position as the actual residual fitting error normalized to the natural residual fitting error:

$$\psi^2 = \frac{(\varepsilon - \varepsilon_\mu)^2}{\varepsilon_\sigma^2} \tag{1}$$

Intuitively, ψ measures in what interval $\varepsilon_\mu \pm \psi \cdot \varepsilon_\sigma$ the error is. The role of this error normalization is also illustrated in Fig. 3.
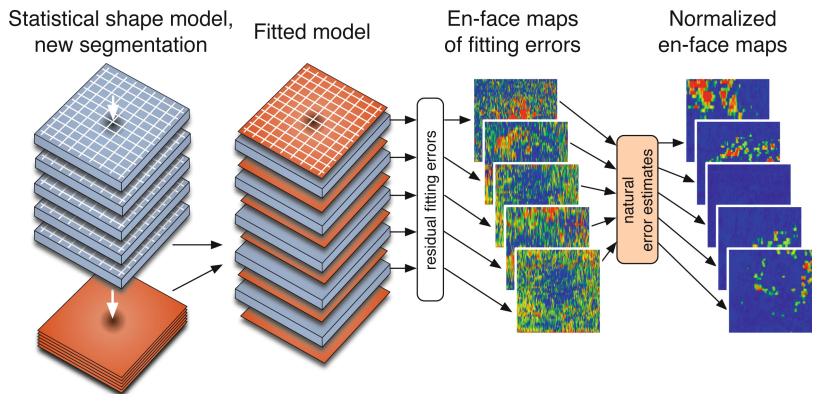
Fig. 3. Statistical shape model fitting: The statistical shape model (grey) is deformed to fit to the surfaces of a new segmentation (red) and the residual error is normalized to build the hinting en-face map for each surface
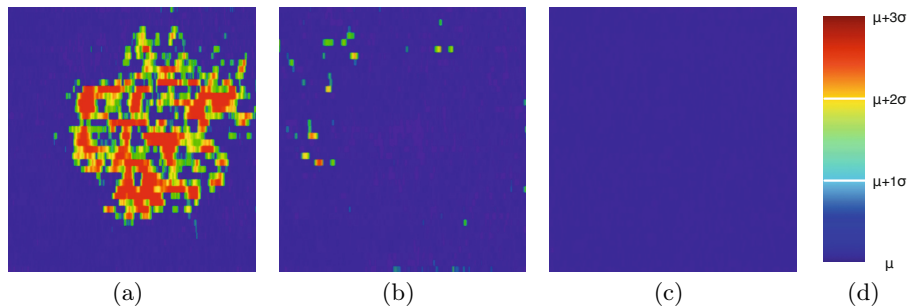
Fig. 4. Thickness hinting en-face maps of the RPE thickness: (a) AMD patient with large drusen, (b) small drusen, (c) healthy RPE, (d) used color transfer function

To present the computed measurements to the diagnostician, an en-face map is built for each layer thickness. Every landmark becomes a pixel in this en-face image. A color transfer function is used to map the values from (1) to a color value. Fig.4 shows en-face maps for two dataset with drusen and a healthy one. The error can of course also be projected back into the OCT volume. Fig.5 shows an example B-scans containing drusen where the cell layers between the segmentation surfaces are colored by the hinting system. The drusen are clearly made visible by the hinting. Note also the abnormally thin cell layer above the drusen to the right of the fovea.

We also implemented a statistical shape model using the full 3D positions of the segmented surfaces. This model is less accurate in detecting small thickness abnormalities, but can reliably visualize abnormalities in the actual shape of the segmented surfaces. Fig.6 shows an example of a patient with an abnormally shaped retina.
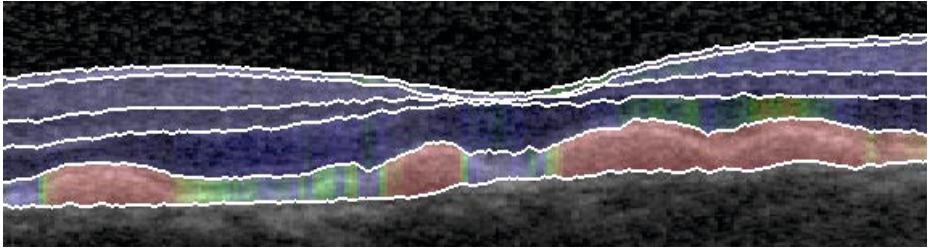
Fig. 5. B-scan with the residual fitting error projected back into the OCT volume

(a)   (b)
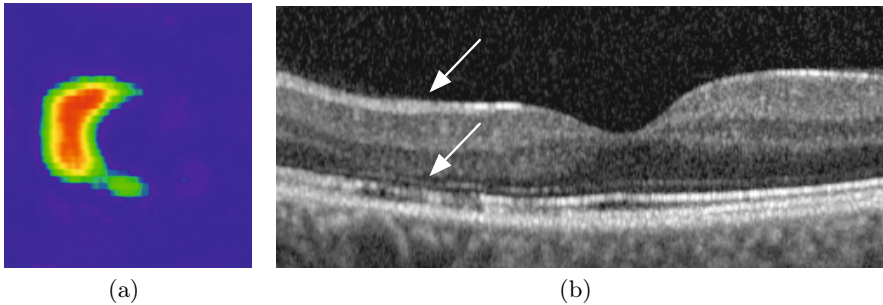
Fig. 6. (a) Shape deformation analysis of the ILM of a patient with a degenerated outer nuclear layer (lower arrow), leading to an abnormally shaped ILM (upper arrow), visible in the corresponding hinting en-face map in (b)

3   Experiments and Results

3.1   Segmentation Evaluation

The accuracy of the hinting system depends greatly on the accuracy of the segmentation and a thorough evaluation was therefore performed. The segmentation algorithm needs to be able to segment healthy datasets as well as datasets with drusen. We therefore evaluated the algorithm on both healthy and AMD datasets. 20 OCT volumes, each having five randomly chosen B-scans manually segmented by two experts, were used in the evaluation. The average manual segmentation from both observers was compared to the result of the automatic segmentation algorithm. As we evaluated the hinting system on drusen, the evaluation of the segmentation focused on the two layers relevant for drusen segmentation: the lower BM and upper IS/OS boundaries.

The upper IS/OS boundary was segmented with a mean unsigned error of $1.69 \pm 1.61 \mu$m. The lower BM boundary was segmented with a mean unsigned error of $2.75 \pm 2.49 \mu$m.

To evaluate the accuracy of the segmentation on actual drusen, 20 datasets containing drusen were used. In every dataset, the lower BM and upper IS/OS boundary of every druse up to a height of 141μm was segmented manually. The

healthy parts of the RPE were not segmented and not used in the evaluation. This allowed us to evaluate the accuracy of the algorithm only on the positions where drusen were present. For the upper IS/OS boundary, the mean unsigned error over all drusen was $5.66 \pm 10.00\mu m$. The mean unsigned error of the lower BM boundary was $4.67 \pm 7.42\mu m$ over all drusen.

While the error for drusen is larger than the error for healthy tissue, it is still relatively small considering the axial pixel resolution of $3.9\mu m$. The result is also comparable to the results by Garvin et al. [1], who reported an unsigned mean error of $3.30 \pm 1.60\mu m$ for the upper IS/OS boundary on healthy datasets. Furthermore, the larger segmentation errors were observed for large drusen that did not have a clear gradient. However, in this case, a larger error is not a problem as the segmentation is still very much outside the range of segmentations of healthy eyes and the drusen are clearly visible in the hinting system.

3.2   Hinting Evaluation

To evaluate the statistical shape model, a leave-one-out test was performed. The mean unsigned error over all landmark points over all datasets was $5.71 \pm 7.86\mu m$.

To evaluate the actual hinting, we implemented a screening system where experts had to decide purely from the en-face map of the RPE thickness if a dataset contained drusen. 20 datasets with drusen were mixed with 20 datasets of healthy eyes. The datasets were presented in random order to each experts and they had to decide for each one whether it contained drusen. One dataset with only a few small drusen was missed by one of the two experts, all other datasets containing drusen were correctly identified by both experts. The combined false positive ratio of the two expert was 10% (4 out of 40). As most datasets contained very large drusen, this test was very easy for the experts. Nevertheless, it shows that the hinting system works well on real data.

As we wanted to know how sensitive the hinting system is, we implemented an evaluation procedure with artificially created small drusen. 50 datasets of 50 healthy subjects were used. When a dataset was presented to the expert, a single druse was created with a probability of 50%. The position of the druse was randomly chosen, as was the height displacement, ranging from just one pixel to 15 pixels in the OCT volume. With an axial resolution of $3.9\mu m$, the resulting height displacement was in the range of 3.9 to $58.5\mu m$. Table 1 shows the results of the evaluation. Most drusen with a height of $85.5\mu m$ were already detected, and all drusen with a height of at least $93.6\mu m$ were detected. The experts spent an average duration of only 7.3 ± 4.8 seconds looking at the en-face map before making a decision. This included loading the segmentation, fitting of the statistical shape model and visualization. As expected, experts had difficulties identifying the smaller drusen, as they often were not sure if a slightly abnormal looking area was healthy and had to guess. In a real clinical setting, the OCT would of course be made visible as well. A click at an abnormal looking area in the hinting map could for example open the OCT volume at that position for the diagnostician to check.

Table 1. Combined results of the two experts identifying drusen from only the hinting en-face map. The column 'AMD' is the result of the test with patients with drusen.

Druse height (μm)	no druse	70.2	74.1	78.0	81.9	85.5	89.70	93.6-136.5	AMD
Identified as drusen	23	0	3	2	3	8	7	49	39
Identified as healthy	129	5	1	4	3	2	1	0	1
Correctly identified (%)	84.9	0	75	33.3	50	80	87.5	100	97.5

4   Conclusion

In this work a new method to quickly detect morphological abnormalities in medical images is proposed. By combining automatic segmentation with a statistical shape model, we were able to visualize and detect most drusen in OCT volumes. With an average time of only 7.3 ± 4.8 seconds spent on each dataset, experts were able to very quickly screen the datasets. The proposed method is completely complementary to the current clinical practice of visual inspection of each individual B-scan of the OCT volume. The accuracy and speed make this method a valuable tool for both large-scale screening systems and the daily clinical practice. Furthermore, it can be applied in virtually any field where automatic segmentation and statistical modeling of anatomy is possible. We are currently working on extending the method to include the analysis of texture information by using a statistical model of appearance.

References

1. Garvin, M.K., Abràmoff, M.D., Wu, X., Russell, S.R., Burns, T.L., Sonka, M.: Automated 3-D intraretinal layer segmentation of macular spectral-domain optical coherence tomography images. IEEE Transactions on Medical Imaging 28(9), 1436–1447 (2009)
2. Song, Q., Wu, X., Liu, Y., Garvin, M., Sonka, M.: Simultaneous searching of globally optimal interacting surfaces with shape priors. In: IEEE International Conference on Computer Vision and Pattern Recognition (2010)
3. 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)
4. Wu, X., Chen, D.Z.: Optimal Net Surface Problems with Applications. In: Widmayer, P., Eidenbenz, S., Triguero, F., Morales, R., Conejo, R., Hennessy, M. (eds.) ICALP 2002. LNCS, vol. 2380, pp. 1029–1042. Springer, Heidelberg (2002)
5. Li, K., Wu, X., Chen, D.Z., 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)
6. Ryan, S.J., Wilkinson, C.P.: Retina. Mosby (2005)

An Invariant Shape Representation Using the Anisotropic Helmholtz Equation

A.A. Joshi[1], S. Ashrafulla[1], D.W. Shattuck[2], H. Damasio[3], and R.M. Leahy[1],[⋆]

[1] Signal & Image Processing Institute, Univ. of Southern California, Los Angeles, CA
[2] Laboratory of Neuro Imaging, Univ. of California, Los Angeles, CA
[3] Brain and Creativity Institute, Univ. of Southern California, Los Angeles, CA

Abstract. Analyzing geometry of sulcal curves on the human cortical surface requires a shape representation invariant to Euclidean motion. We present a novel shape representation that characterizes the shape of a curve in terms of a coordinate system based on the eigensystem of the anisotropic Helmholtz equation. This representation has many desirable properties: stability, uniqueness and invariance to scaling and isometric transformation. Under this representation, we can find a point-wise shape distance between curves as well as a bijective smooth point-to-point correspondence. When the curves are sampled irregularly, we also present a fast and accurate computational method for solving the eigensystem using a finite element formulation. This shape representation is used to find symmetries between corresponding sulcal shapes between cortical hemispheres. For this purpose, we automatically generate 26 sulcal curves for 24 subject brains and then compute their invariant shape representation. Left-right sulcal shape symmetry as measured by the shape representation's metric demonstrates the utility of the presented invariant representation for shape analysis of the cortical folding pattern.

1 Introduction

The human cerebral cortex is a highly convoluted sheet with rich and detailed folding patterns. Sulci are fissures in the cortical surface which are used frequently as anatomical landmarks. The geometry of these cortical landmarks is used for registration as well as the study of disease progression [10], aging [12] and brain asymmetry [1]. However, these approaches do not use the shapes of sulci but instead features such as length, depth and 3D location.

Quantification, matching, and classification of the shape of curves is a challenging problem with a long history. Spectral graphs [3] use graph theory to attempt to match two curves. In addition, geometric features such as areas of enclosed regions [21] have been used for curve representation. Recent methods use the distributions of distances from all points on a curve to a reference point; the most popular is the shape context [9].

PDE based models such as elastic matching [19] LDDMM [5], conformal representation [16], bimorphisms [20] use locally smooth deformation models for curve registration. In general, these methods find transform parameters and then calculate a distance on the final fit [2]. However, it is unclear whether they capture both the local and the global features of a curve.

In order to address these issues in the context of sulcal pattern analysis, we present a novel model for shape analysis of 1D curves based on an extension of the Global Point Signature (GPS) [13,14]. The GPS representation for 2D surfaces uses the eigensystem of the Laplace-Beltrami operator. We first review our method for labeling the cortical surface automatically and for generating sulci on a population of cortical surfaces. We next present a novel invariant 1D curve representation based on the extension of GPS to 1D curves. For consistency of terminology with the shape analysis literature, we refer to our novel representation as the GPS representation of curves. 1D curves are represented using the eigensystem of an anisotropic Helmholtz equation where curvature is used as the anisotropy term. This representation also defines a metric in the embedded space of the representation - admitting a local measure of shape distance for curve matching. Finally, we present the results of applying our methodology towards analysis of sulcal shape symmetry between left and right brain hemispheres.

2   Sulci Generation

We briefly review our method for automatic generation of sulci on a subject's cortical surface, described in more detail in [7]. We assume as input a triangulated mesh approximating the cortical surface. We use BrainSuite [17] to extract the cortical surface meshes from T1-weighted MRI volumes for an atlas and for the subject. We then identify sulcal landmarks on the cortex automatically. We compute a one-to-one correspondence between the atlas surface and the subject surface in two stages: (i) the surface of each cortical hemisphere is parameterized to a unit square, and (ii) a vector field is found within this parameterization that aligns curvature of the atlas surface to curvature of the subject surface. We parameterize the cortical surfaces by modeling them as an elastic sheet and then solving the associated linear elastic equilibrium equation using finite elements. We constrain the corpus callosum to lie uniformly on the boundary of the unit square. The elastic energy minimization yields flat maps of each cortical hemisphere to a plane (Fig. 1). Multiresolution representations of curvature for the subject and atlas are calculated and then aligned by minimizing a cost function with elastic energy as a regularizer. This step reparameterizes the cortical hemisphere surfaces, establishing a one-to-one point correspondence between subject and atlas surfaces.

For this study we registered $N = 24$ T1-weighted MRI volumes. For the atlas, a set of 26 sulcal curves per hemisphere were traced interactively in BrainSuite [17] using the protocol described in [11]. Using the point correspondence

established above, these sulci are transferred to the subject surface. The locations of the estimates are subsequently refined using geodesic curvature flow on the cortical surface as described in [7]. The refinement uses a level set based formulation of flow on non-flat surfaces with the sulci as the zero level sets. The resulting PDE is discretized on a triangulated mesh using finite elements. After this refinement, we have 26 sulci on each of $N = 24$ cortical surfaces. We note that other methods for sulcal set generation [18] can also be used.
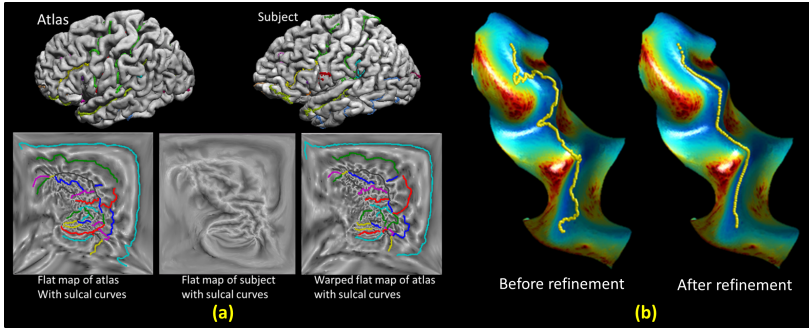
Fig. 1. (a) Automatic atlas to subject registration and parameterization of cortical surfaces and sulcal curves; and (b) geodesic curvature flow refinement of sulcal curves

3   Shape Representation Using GPS Representation

In this section we introduce a coordinate system to represent a 1D curve. Spectral theory provides the basis to study the eigenspectrum of the sulcal curves. Motivated by spectral theory and corresponding work on 2D surfaces [14,13], we model the 1D curves as inhomogeneous vibrating strings. Their harmonic behavior is governed by the 1D Helmholtz equation. To characterize the shape of the curve C, we use its curvature $\kappa(s)$ to introduce anisotropy into the governing equation:

$$\begin{cases} \frac{\partial}{\partial s}\kappa(s)\frac{\partial}{\partial s}\Phi_i(s) & = \lambda_i\Phi_i(s) \\ \Phi_i(s)|_{\partial C} & = 0 \end{cases}, \forall s \in C \tag{1}$$

where ∂C is the set of the endpoints of the curve C. Denote the eigenfunctions of this equation by Φ_i with eigenvalues λ_i ordered by magnitude. We define the embedding manifold in the spectral domain by the map:

$GPS(p) = \left(\frac{1}{\sqrt{\lambda_1}}\Phi_1(p), \frac{1}{\sqrt{\lambda_2}}\Phi_2(p), \frac{1}{\sqrt{\lambda_3}}\Phi_3(p), \dots\right).$

Each point of the curve is embedded into an infinite dimensional space. We cannot use the 1D Laplacian directly for this purpose because 1D shapes always have a trivial intrinsic geometry. However, due to the fundamental theorem of curves (two unit-speed plane curves that have the same curvature and torsion differ only by a rigid transformation), curvature and torsion define the curve uniquely up to a rigid transformation. Furthermore, the curve can be recovered
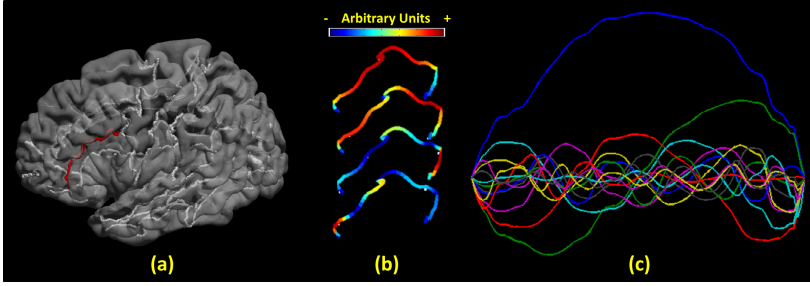
Fig. 2. (a) Inferior frontal sulcus highlighted in red; (b) first four color coded GPS coordinates; (c) GPS representation of a sulcus plotted from end to end

from the embedding by first recovering the curvature and torsion, and then using the Frenet-Serre formulas [4]. For curves in 3D space, this requires curvature and torsion. The embedding defined above is based on curvature alone because the sulcal curves analyzed in this paper had negligible torsion. The following properties also apply to this representation:

1. GPS coordinates are isometry invariant as they depend only on derivatives and curvature, which are ony dependent on shape.
2. Scaling a 1D curve manifold equally scales curvature. Therefore, normalizing the eigenvalue produces scale invariance (as well as position invariance, as mentioned earlier).
3. Shape changes result in continuous changes in a curve's spectrum. Consequently the representation presented here is robust.
4. In the embedding space, the inner product is given by the Green's function due to the identity: $G(x_1, x_2) = \sum_i \frac{\Phi_i(x_1)\Phi_i(x_2)}{\lambda_i}$. Thus the GPS representation encodes local and global shape information. Additionally, the metric is Euclidean.

An example of this representation for a sulcal curve is shown in Fig. 2.

4   Discretization Using Finite Element Method

Sulcal curves are often sampled non-uniformly, so we use a finite element method to discretize the eigenvalue problem in Eq. 1. Let $\Phi(s) = \sum_i \phi_i e_i(s)$ be an eigenfunction and $\eta(s) = \sum_i \eta_i e_i(s)$ be a 'test function' represented as weighted sums of linear elements. The eigenvalue problem from Eq. 1 is:

$$\left(\frac{\partial}{\partial s} \kappa(s) \frac{\partial}{\partial s} \right) \Phi = \lambda \Phi$$

$$\implies \int \left(\frac{\partial}{\partial s} \kappa(s) \frac{\partial}{\partial s} \Phi(s) \right) \eta(s) ds = \lambda \int \Phi(s) \eta(s) ds$$

$$\implies \int \kappa(s) \frac{\partial}{\partial s} \Phi(s) \frac{\partial}{\partial s} \eta(s) ds = \lambda \int \Phi(s) \eta(s) ds$$

where the latter follows using integration by parts. Substituting the finite element model we get:

$$\sum_i \sum_j \phi_i \eta_j \kappa_{ij} \int \frac{\partial}{\partial s} e_i(s) \frac{\partial}{\partial s} e_j(s) ds = \lambda \sum_i \sum_j \phi_i \eta_j \int e_i(s) e_j(s) ds$$

$$\kappa S \phi = \lambda M \phi \tag{2}$$

where κ_{ij} represents $(\kappa_i + \kappa_j)/2$: the average of curvatures calculated at points i and j.

For the 1D case with linear elements, the element-wise mass matrix is given by $M_{el} = \begin{bmatrix} (\kappa_{ij} d_{ij})/3 & (\kappa_{ij} d_{ij})/6 \\ (\kappa_{ij} d_{ij})/6 & (\kappa_{ij} d_{ij})/3 \end{bmatrix}$ for element el corresponding to the edge between nodes i and j. Similarly, the element-wise stiffness matrix is given by $S_{el} = \begin{bmatrix} 1/d_{ij} & -1/d_{ij} \\ -1/d_{ij} & 1/d_{ij} \end{bmatrix}$ using linear finite elements [15].

The matrix equation in Eq. 2 is a generalized sparse eigenvalue problem that can be solved using standard methods, such as the QZ method that is a part of the Matlab function `eigs`. The point-wise curvature of the curve κ_i is computed using the Frenet frame [4].

5   Shape Matching

In brain image analysis, a matching technique is required to analyze sulcal variation across a population. In this section, we describe a method for finding such a matching using GPS coordinates. Later, we match left vs right hemispherical sulci to investigate asymmetry between hemispheres.

Let GPS_1 and GPS_2 denote the GPS coordinates for the two sulcal sets. Our goal is to find a reparameterization function ψ such that the matching energy $E(\psi)$ is minimized.

$$E(\psi) = \int ||(GPS_1(s) - GPS_2(s + \psi(s)))||^2 ds \tag{3}$$

where ψ is represented in terms of b-spline basis functions [15]. Minimization of the cost function results in a 1-1 point correspondence between the two curves (Fig 3). Once the optimal ψ is found, the local shape difference at point s is given by $||(GPS_1(s) - GPS_2(s + \psi(s)))||$.

For the purpose of mapping symmetry, we compute (a) the point-wise GPS distance between corresponding sulci from one hemisphere to the other, for all subjects; (b) the point-wise GPS distance between corresponding sulci for the same hemisphere in two different subjects. We define a measure of symmetry $Symm = -log(\frac{mean(a)}{mean(b)})$. The measure $Symm$ ranges from 0 to ∞. We then use
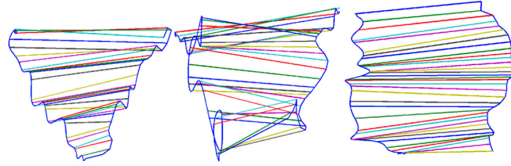
Fig. 3. Three representative sulci from left and right hemispheres and the point correspondence between them

a non-parametric Mann–Whitney–Wilcoxon test between statistics (a) and (b) at $\alpha = 0.05$, correcting for multiple comparisons, with the false discovery rate (FDR).

6 Results

We performed symmetry detection on data from 24 subjects, divided into two cohorts of 12 subjects [11]. The first cohort was scanned at the Dornsife Cognitive Neuroscience Imaging Center using a 3T Siemens MAGNETOM Trio scanner. The second cohort was scanned at the University of Iowa using thin-cut MR coronal images obtained in a General Electric Signa Scanner operating at 1.5 Tesla. We applied the BrainSuite surface extraction sequence followed by sulcal set generation as outlined in Sec. 2. This produced 24x2 cortical surface hemisphere representations with 26 sulci each. The sulci were denoised by fitting a 12th order polynomial with the degree of the polynomial was selected using L-curve analysis and selecting the maximum degree necessary for all curves. Next, the GPS coordinate representation was formed as described in Sec. 3 and Sec. 4. The symmetry between the sulci was then estimated using the method in Sec 5. The results of the symmetry mapping are shown in Fig. 4. It is interesting to note that the post- and pre-central sulci, together with the posterior segment of the superior temporal, the transverse temporal, the middle temporal and the inferior occipital sulci in the dorso-lateral view, show the maximal amount of left/right asymmetry; on the mesial view the collateral, the supraorbital, the occipito-parietal and long stretches of the cingulate sulci are also extremely asymmetric. It is not surprising to see the cingulate sulcus (visible in the depth of the mesial view) show a great extent of relative symmetry. Reports of brain asymmetry usually focus on the Sylvian fissure but our data suggest that other asymmetries may be worth investigating to determine if they are indeed comparable to the well confirmed Sylvian fissure asymmetry.

It would be interesting to apply other shape representation approaches for the problem of finding sulcal shape symmetry. In this work, we tried a simple affine curve matching approach [6] that finds an optimal affine transform to minimize distance between curves but but we did not find significant asymmetries with this affine approach.
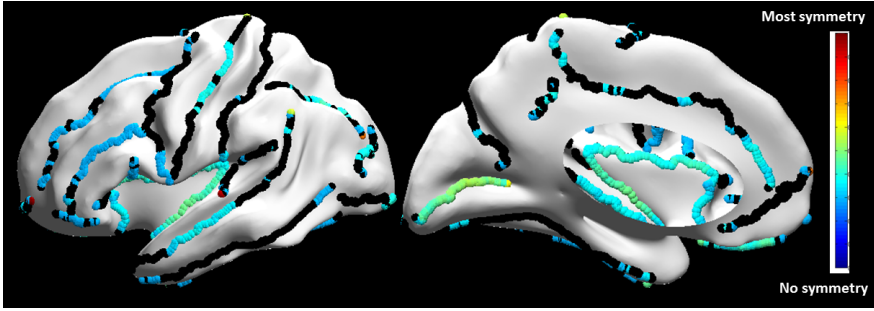
Fig. 4. Shape symmetry measure of the sulci plotted on a smooth representation of an individual cortical surface. The black regions on the curves indicate that a significant symmetry was not found for those points.

7 Discussion and Conclusion

We have presented a novel invariant shape representation using the eigensystem of the anisotropic Helmholtz equation. This representation also has an interesting physical interpretation in terms of vibrating strings. Because our representation depends only on shape and not on the Euclidean embedding of the shape, it is invariant to Euclidean transformations. As opposed to surface-based measures [8,5] in which sulcal shape differences are confounded by the shape of the cortical sheet surrounding the sulcal fold, we model sulci as curves as opposed to folds on surfaces. The invariant representation therefore provides information that is complementary to surface-based shape analysis. The properties listed Sec 3 make the presented GPS representation for curves an attractive alternative over the existing methods [20,16,9,19], although a thorough comparison is still required.

One potential drawback of our method is that errors in automatically generated sulci can lead to inaccurate input when generating the GPS representation. We are in the process of validating the sulcal generation method in a more extensive manner on a larger data-set; initial validation is promising. It is important to note that, if required, the BrainSuite software allows for semi-automatic interactive corrections of the sulci to reduce inaccuracies.

This model has a variety of potential applications in computer vision as well as brain image analysis. Many of the existing methods for brain morphometry focus on point-wise features such as 3D location, curvature, thickness, deformation, and image intensity. Conversely, the framework we have presented directly captures the geometric shape of the folding pattern. As a result, we can study the cortical folding pattern quantitatively for a variety of neuro-developmental conditions (e.g. autism) and other neurological conditions characterized by changes in sulcal patterns.

References

1. Blanton, R., Levitt, J., Thompson, P., Narr, K., Capetillo-Cunliffe, L., Nobel, A., Singerman, J., McCracken, J., Toga, A.: Mapping cortical asymmetry and complexity patterns in normal children. Psychiatry Research: Neuroimaging 107(1), 29–43 (2001)
2. Buchin, K., Buchin, M., Wang, Y.: Exact algorithms for partial curve matching via the Fréchet distance. In: Proc. of the 20th Ann. ACM-SIAM Symp. on Discrete Algorithms, pp. 645–654 (2009)
3. Carcassoni, M., Hancock, E.: Spectral correspondence for point pattern matching. Pattern Recognition 36(1), 193–204 (2003)
4. Do Carmo, M.: Differential Geometry of Curves and Surfaces. Prentice-Hall (1976)
5. Glaunès, J., Qiu, A., Miller, M., Younes, L.: Large deformation diffeomorphic metric curve mapping. International Journal of Computer Vision 80, 317–336 (2008)
6. Gope, C., Kehtarnavaz, N., Hillman, G., Würsig, B.: An affine invariant curve matching method for photo-identification of marine mammals. Pattern Recognition 38(1), 125–132 (2005)
7. Joshi, A.A., Shattuck, D.W., Damasio, H., Leahy, R.M.: Geodesic curvature flow on surfaces for automatic sulcal delineation. In: Proc. ISBI (in press, 2012)
8. Lai, R., Shi, Y., Sicotte, N., Toga, A.: Automated corpus callosum extraction via laplace-beltrami nodal parcellation and intrinsic geodesic curvature flows on surfaces. UCLA Computational and Applied Mathematics Reports
9. Mori, G., Belongie, S., Malik, J.: Efficient Shape Matching Using Shape Contexts. IEEE Trans. on PAMI 27(11), 1832–1837 (2005)
10. Narr, K., Thompson, P., Sharma, T., Moussai, J., Zoumalan, C., Rayman, J., Toga, A.: Three-dimensional mapping of gyral shape and cortical surface asymmetries in schizophrenia: gender effects. Am. J. Psychiatry 158(2), 244–255 (2001)
11. Pantazis, D., Joshi, A., Jiang, J., Shattuck, D., Bernstein, L., Damasio, H., Leahy, R.: Comparison of landmark-based and automatic methods for cortical surface registration. Neuroimage 49(3), 2479–2493 (2010)
12. Rettmann, M.E., Prince, J.L., Resnick, S.M.: Analysis of sulcal shape changes associated with aging. In: Proc. Human Brain Mapping, NY, June 18-22 (2003)
13. Reuter, M.: Hierarchical Shape Segmentation and Registration via Topological Features of Laplace-Beltrami Eigenfunctions. IJCV 89(2), 287–308 (2010)
14. Rustamov, R.: Laplace-beltrami eigenfunctions for deformation invariant shape representation. In: Proc. of the 5th Euro. Symp. on Geom. Proc., pp. 225–233 (2007)
15. Sadiku, M.: Numerical techniques in electromagnetics. CRC (2000)
16. Sharon, E., Mumford, D.: 2D-shape analysis using conformal mapping. International Journal of Computer Vision 70(1), 55–75 (2006)
17. Shattuck, D.W., Leahy, R.M.: Brainsuite: An automated cortical surface identification tool. Medical Image Analysis 8(2), 129–142 (2002)
18. Shi, Y., Tu, Z., Reiss, A., Dutton, R., Lee, A., Galaburda, A., Dinov, I., Thompson, P., Toga, A.: Joint sulcal detection on cortical surfaces with graphical models and boosted priors. IEEE Transactions on Medical Imaging 28(3), 361–373 (2009)
19. Srivastava, A., Klassen, E., Jo, S.H., Jermyn, I.H.: Shape analysis of elastic curves in euclidean spaces. IEEE Trans. on PAMI 33(7), 1415–1428 (2011)
20. Tagare, H.D., O'shea, D., Groisser, D.: Non-rigid shape comparison of plane curves in images. J. Math. Imaging Vis. 16(1), 57–68 (2002)
21. Xu, C., Liu, J., Tang, X.: 2D Shape Matching by Contour Flexibility. IEEE Trans. on PAMI 31(1), 180–186 (2009)

Phase Contrast Image Restoration
via Dictionary Representation of Diffraction
Patterns

Hang Su[1,3], Zhaozheng Yin[2], Takeo Kanade[3], and Seungil Huh[3]

[1] Department of EE, Shanghai Jiaotong University
[2] Department of CS, Missouri University of Science and Technology
[3] The Robotics Institute, Carnegie Mellon University[⋆]

Abstract. The restoration of microscopy images makes the segmentation and detection of cells easier and more reliable, which facilitates automated cell tracking and cell behavior analysis. In this paper, the authors analyze the image formation process of phase contrast images and propose an image restoration method based on the dictionary representation of diffraction patterns. By formulating and solving a min-ℓ_1 optimization problem, each pixel is restored into a feature vector corresponding to the dictionary representation. Cells in the images are then segmented by the feature vector clustering. In addition to segmentation, since the feature vectors capture the information on the phase retardation caused by cells, they can be used for cell stage classification between intermitotic and mitotic/apoptotic stages. Experiments on three image sequences demonstrate that the dictionary-based restoration method can restore phase contrast images containing cells with different optical natures and provide promising results on cell stage classification.

1 Introduction

Computer-aided image analysis of phase contrast microscopy [1] has attracted increasing attention since it enables long-term monitoring of the proliferation and migration processes of live cells. Among the tasks of microscopy cell image analysis, cell detection and segmentation is one of the most fundamental components in that various analyses can be performed based on it. Cell detection and segmentation in phase contrast microscopy is challenged by clustered cells, cell shape deformation, and image artifacts such as bright halos and shade-off.

The common techniques employed for cell segmentation include thresholding [2], edge detection, and morphological operations [3]. These methods often fail when the contrast between cells and background is low. Another group of segmentation algorithms that are based on intensity gradient of images, namely, watershed [4], active contours [5], and level set [6], are sensitive to the initializations and local noisy gradients. To address these drawbacks, a restoration-based segmentation was recently proposed [7]. The method models the image formation process of phase contrast microscope to restore phase retardation caused by

---

cells, based on which cells are detected. However, this method fails to segment cells when they are bright in phase contrast microscopy images, e.g., mitotic or apoptotic cells, because the model assumes that the phase retardation caused by cells is small, which is not valid when cells become thick and thus appear bright in phase contrast microscopy.

In this paper, we revisit the phase contrast imaging model in [7], and propose a novel restoration algorithm based on dictionary representation of diffraction patterns, which can restore phase contrast images with various phase retardations. The proposed dictionary-based method restores a feature vector corresponding to diffraction patterns for each pixel. After the image restoration, high quality segmentation is achieved by clustering the feature vectors. Furthermore, since the restored feature vectors embed phase retardation information, cells can be classified between different stages, particularly between intermitosis and mitosis/apoptosis(dead).

2   Methodology

2.1   The Image Formation of Phase Contrast Microscope

The phase contrast microscope converts the phase difference in light passing through the transparent specimen to brightness changes in the image [1]. The wavefront of the illuminating beam is divided into two components after passing through the specimen. The primary component is the surround wave (S wave) that passes through or around the specimen without interacting with it. The other component is the diffracted wave (D wave) that is scattered by the specimen. These two waves undergo interference and produce a resultant particle wave (P wave). The cells can be observed only when the amplitudes of P wave and S wave are significantly different.

In [7], the surround wave l_S and the diffracted wave l_D are derived as:

$$l_S = i\zeta_p A e^{i\beta}, \text{ and } l_D = \zeta_c A e^{i(\beta+\theta(x))} + (i\zeta_p - 1)\zeta_c A e^{i(\beta+\theta(x))} \cdot airy(r), \quad (1)$$

where A and β are the amplitude and phase of the incident light, respectively; ζ_c and ζ_p are the amplitude attenuation factors caused by cells and phase ring, respectively; $\theta(x)$ is the phase retardation caused by the specimen at location x; and $airy(r)$ is an obscured Airy pattern (diffraction pattern with a bright region in the center surrounded by a series of concentric dark and bright rings [7]). The particle wave l_P is calculated as $l_P = l_S + l_D$.

During the imaging model derivation in [7], the exponential terms in equation (1) are approximated using $e^{i\theta(x)} \approx 1 + i\theta(x)$. Note that this approximation is valid only when the phase retardation $\theta(x)$ is close to zero. The assumption is apparently not applicable to general cases since $\theta(x)$, which is a function of the refractive indices and the thickness of cells, often vary along with different cell types and stages; more formally, the phase retardation θ can be calculated as:

$$\theta = \frac{2\pi}{\lambda}(n_1 - n_2)t, \quad (2)$$

where λ denotes the wavelength of the incident light; t is the thickness of the cell; and n_1 and n_2 denote the refractive indices of the cell and medium, respectively.

2.2   Dictionary Representation of Diffraction Patterns

In this paper, we propose a generalized imaging model by approximating the term $e^{i\theta(x)}$ in Eq. (1) using a linear combination of $\{e^{i\theta_m}\}$:

$$e^{i\theta(x)} \approx \sum_{m=0}^{M-1} \psi_m(x) e^{i\theta_m}, \quad s.t. \sum_{m=0}^{M-1} \psi_m(x) = 1 \ \ and \ \ \psi_m(x) \geqslant 0. \tag{3}$$

where $\{\theta_m\} = \{0, \frac{2\pi}{M}, \cdots, \frac{2m\pi}{M}, \cdots, \frac{2(M-1)\pi}{M}\}$. We impose the nonnegative constraint because the solution would not be unique without it as $\psi_m(x) e^{i\theta_m} = -\psi_m(x) e^{i(\theta_m + \pi)}$. Moreover, the nonnegative constraint removes the absolute operator of ℓ_1 norm, allowing the problem to be solved in a standard manner.

The intensity of the observed image g is calculated as:

$$\begin{aligned} g = \|l_p\|^2 &= (l_S + l_D) \cdot (l_S + l_D)^* \\ &= A^2 \left\{ \begin{array}{l} \zeta_p{}^2 + \zeta_c{}^2 - 2\zeta_c^2 \cdot airy(r) + (\zeta_p^2 + 1)\zeta_c^2 \cdot (airy(r))^2 + \\ \zeta_p^2 \zeta_c (e^{-i\theta(x)} + e^{i\theta(x)}) \cdot airy(r) + i\zeta_p \zeta_c (e^{-i\theta(x)} - e^{i\theta(x)}) \end{array} \right\}. \end{aligned} \tag{4}$$

Substituting the exponential terms in Eq. (4) with Eq. (3) yields a linear representation of the observed image:

$$\begin{aligned} g &= A^2 \left\{ \begin{array}{l} \zeta_p{}^2 + \zeta_c{}^2 - 2\zeta_c^2 \cdot airy(r) + (\zeta_p^2 + 1)\zeta_c^2 \cdot (airy(r))^2 + \\ \sum_{m=0}^{M-1} \psi_m(x) \Big((2\zeta_p \zeta_c \sin\theta_m \cdot \delta(r) + 2\zeta_p^2 \zeta_c \cos\theta_m \cdot airy(r) \Big) \end{array} \right\}. \\ &= C + D \sum_{m=0}^{M-1} \psi_m(x) \Big(\sin\theta_m \cdot \delta(r) + \zeta_p \cos\theta_m \cdot airy(r) \Big) \end{aligned} \tag{5}$$

where $\delta(\cdot)$ is a Dirac delta function, C is a constant that indicates the items unrelated to the feature vector $\psi_m(x)$, and D is also a constant. C can be eliminated by flat-field correction [1,7] and thus we ignore it for simplicity. Hence,

$$\begin{aligned} g &\propto \sum_{m=0}^{M-1} \psi_m(x) \Big(\sin\theta_m \cdot \delta(r) + \zeta_p \cos\theta_m \cdot airy(r) \Big) \\ &\triangleq \sum_{m=0}^{M-1} \psi_m(x) PSF(\theta_m), \end{aligned} \tag{6}$$

where PSF denotes the point spread function; i.e., $PSF(\theta_m)$ represents the diffraction pattern with phase retardation θ_m. In our experiments, ζ_p was set between 0.4 and 0.5 based on the information of microscope we used.

We discretize $PSF(\theta_m)$ into a $(2T+1) \times (2T+1)$ kernel, (u, v) element of which is denoted by $\widehat{PSF}(\theta_m, u, v)$. Then, from Eq. (6), the imaging model of (i, j) pixel of g is also discretized as:

$$g(i,j) = \sum_{m=0}^{M-1} \sum_{u=1}^{2T+1} \sum_{v=1}^{2T+1} \psi_m(i+u-T-1, j+v-T-1)\widehat{PSF}(\theta_m, u, v). \quad (7)$$

We define Ψ_m as the vectorized representation of the matrix $\{\psi_m(i,j)\}$ and $h_m(i,j)$ as the vector obtained by vectorizing the sparse matrix whose $(i-T : i+T, j-T : j+T)$ submatrix is $\{\widehat{PSF}(\theta_m, u, v)\}$ and the other elements are zero. Then, Eq. (7) is simplified into:

$$g(i,j) = \sum_{m=0}^{M-1} h_m(i,j)^T \Psi_m, \quad (8)$$

and thus the vectorized form of the phase contrast microscopy image $\{g(i,j)\}$ can be modeled as:

$$g = \sum_{m=0}^{M-1} H_m \Psi_m, \quad s.t. \ \Psi_m \geqslant 0 \quad (9)$$

where H_m is the matrix obtained by stacking up row vectors $\{h_m(i,j)^T\}$ in order. Note that when g consists of P pixels in total, H_m is a $P \times P$ matrix, each row of which contains only $(2T+1) \times (2T+1)$ non-zero elements.

2.3   The Restoration of the Phase Contrast Images

Based on the theory of sparse representation [8], we formulate the following optimization problem to restore the feature vector from Eq. (9):

$$min \sum_{k=0}^{N-1}\{\|\Psi_{m_k}\|_1 + w_s \Psi_{m_k}^T \mathbf{L} \Psi_{m_k}\} \ s.t. \begin{cases} \|g - \sum_{k=0}^{N-1} H_{m_k} \Psi_{m_k}\|_2 < \varepsilon \\ \Psi_{m_k} \geqslant 0 \end{cases} \quad (10)$$

where $\mathbf{L}$ is a Laplacian matrix defining the similarity between spatial pixel neighbors [7]; w_s is the weight determining the spatial smoothness, which was set between 0.2 to 0.4 in our experiments; and, N is the number of representative retardations, the optimal value of which can vary with the cell type and property.

We propose an iterative optimization algorithm to solve this min-ℓ_1 optimization problem since it is known that there is no closed-form solution for such a problem. We first search the best-matching N bases in the dictionary $\{H_m\}$ with the matching pursuit algorithm [8], and then utilize a non-negative multiplicative updating method [9] to obtain the nonnegative feature vectors $\{\Psi_{m_k}\}$. The procedure is described in Algorithm 1.

Solving Eq. (10) yields the best set of $\{\Psi_{m_1}, \cdots, \Psi_{m_k}\}$, which means that each pixel is restored as a feature vector. We apply K-means clustering on the feature vectors to segment cells in images.

Algorithm 1. Optimization Algorithm for the Image Restoration

---

Input g: vectorized image, $\{H_m\}$ dictionary matrices,
$\quad$ N: number of representative retardations, th: threshold for residual evaluation.
Output $\{\Psi_{m_k}\}$: feature representation vectors.
Initialization $R_g^0 \leftarrow g$; $P \leftarrow$ size of g; $R_g^0 \leftarrow$ initial residual error.
for $k = 0 \rightarrow N - 1$ **do**
$\quad$ //*Search for the best matching basis H_{m_k} in dictionary $\{H_m\}$*
$\quad$ **for** $m = 0 \rightarrow M - 1$ **do**
$\quad\quad$ Compute inner product: $\rho_m(i) = \left\langle H_m(:,i), R_g^k \right\rangle, \forall i \in \{0,\cdots,P-1\}$
$\quad\quad$ **if** m is equal to zero **then** $\rho_k \leftarrow \rho_m$, $H_{m_k} \leftarrow H_m$.
$\quad\quad$ **else if** $\|\rho_m\|_0 > \|\rho_k\|_0$ **then**
$\quad\quad\quad$ $\rho_k \leftarrow \rho_m$, $H_{m_k} \leftarrow H_m$.
$\quad\quad$ **end if**
$\quad$ **end for**
$\quad$ //*Calculate the feature vector Ψ_{m_k}*
$\quad$ $g'(i) = 0, \quad \forall i \in \{0,\cdots,P-1\}$ s.t. $\rho_k(i) \leq th$.
$\quad$ $g'(i) \leftarrow R_g^k(i), \quad \forall i \in \{0,\cdots,P-1\}$ s.t. $\rho_k(i) > th$. //*Assign the relevant elements*
$\quad$ Formulate a subproblem from Eq. (10):
$\quad\quad$ $min\|\Psi_{m_k}\|_1 + w_s\Psi_{m_k}{}^T\mathbf{L}\Psi_{m_k} \quad s.t. \quad \|g' - H_{m_k}\Psi_{m_k}\|_2 < \varepsilon \ and \ \Psi_{m_k} \geqslant 0.$
$\quad$ Obtain the feature vector Ψ_{m_k} by solving this problem with the method [9]
$\quad$ $R_g^{k+1} \leftarrow R_g^k - H_{m_k}\Psi_{m_k}$ //*Update the residual error*
end for

---

3 Experiments and Discussions

Data. The proposed approach was tested on three different sets of phase contrast images of 1040×1392 pixels. The specifications of the datasets are summarized in **Table** 1.

Table 1. Specifications of the Datasets

	Frame number	Cell type	Cell number per image	Cell stages
Seq1	500	bovine aortic endothelial cell	500 to 800+	intermitosis/mitosis
Seq2	600	muscle stem of a progeroid	50 to 300+	intermitosis/mitosis
Seq3	500	C2C12 myoblastic stem cell	300+	intermitosis/apoptosis(dead)

Parameters. Our algorithms involve three parameters: the dictionary size M, the number of representative phase retardations N, and number of classes for clustering K. We determined these parameters by investigating 10 frames for each sequence, which were uniformly sampled. For setting M, we plotted M versus the average residual error in Eq. (3) and set M to be 15 where the residual error levels off as shown in Fig. 1(a). We set N to be 3 for dark cells, bright cells, and background. In addition to these three categories, we took into account two more categories, the boundary between either dark cells or bright cells and background, setting k to be 5, as shown in Figs. 1(b) and (c).
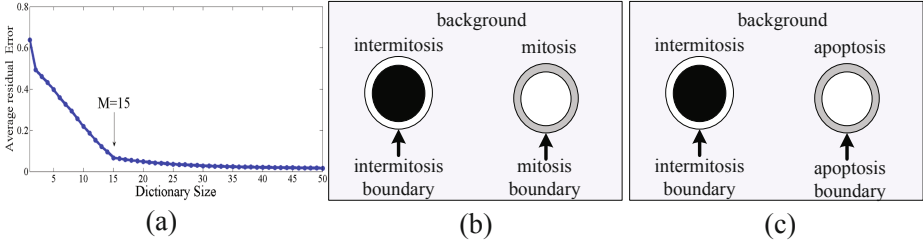
Fig. 1. (a) The average residual error decreases as M increases and it levels off when M is around 15. (b,c) Three representative retardations for intermitosis, mitosis/apoptosis(dead), and background; and five clustering classes for inner intermitosis, intermitosis boundary, inner mitosis/apoptosis, mitosis/apoptosis boundary, and background.
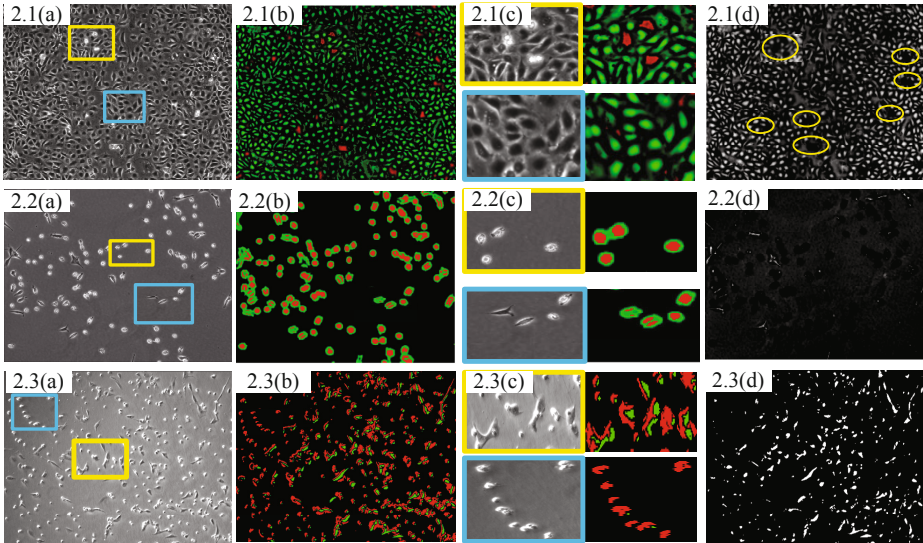
Fig. 2. The sample results of cell segmentation. Top, middle, and bottom rows show the results on Seq1, Seq2, and Seq3, respectively. (a) Original phase-contrast images, (b-c) segmentation results of the proposed method, (d) segmentation results of the previous method [7]. The intermitotic cells are marked with green color and the mitotic/apoptotic(dead) cells with red color. Yellow ellipses in Fig. 2.1(d) indicate missed mitotic cells. Figs. 2.2(d) and 2.3(d) also demonstrate that the previous method fails to detect most of mitotic/apoptotic(dead) cells.

Segmentation: Our dictionary-based approach achieved high quality segmentations as can be seen in Fig. 2. The method well detected bright cells, which undergo mitosis or apoptosis. On the other hand, the previous method [7] failed

Table 2. Quantitative Results on Cell Segmentation

	Seq.1		Seq.2		Seq.3	
	Proposed	Previous	Proposed	Previous	Proposed	Previous
Average Precision	93.1%	87.8%	95.3%	34.8%	94.5%	31.4%
Average Recall	90.2%	85.8%	92.6%	15.7%	91.6%	34.7%
Average F_{score}	0.9154	0.8676	0.9412	0.1749	0.9302	0.3212

Table 3. Quantitative Results on Cell Stage Classification

Seq.1		Seq.2		Seq.3	
Precision	Recall	Precision	Recall	Precision	Recall
71.3%	98.6%	78.3%	97.9%	67.8%	98.3%

to segment mitotic cells in Seq1 and Seq2 as well as apoptotic cells in Seq3, as shown in column (d) of Fig. 2. This result clearly demonstrates that the assumption of the previous method on the phase retardation being close to zero is not valid for bright cells.

Classification: The proposed method is able to not only segment cells, but also classify them among different stages. Fig. 2.1-2.(c) show zoom-in details on the classification between intermitotic cells (green) and mitotic/apoptotic(dead) cells (red). In Fig 2.3.(c), as cell death proceeds, more cells appear red on the results.

Evaluation: In order to evaluate our method quantitatively, we manually labeled every 50th image in Seq1 (4125 annotated cells, 9.1×10^5 cell pixels), every 75th image in Seq2 (1459 annotated cells, 5.1×10^5 cells pixels), and every 100th image in Seq3 (3915 annotated cells, 1.32×10^6 cell pixels). During the test, we skipped the frames used for training. We measured performance in terms of *precision*, *recall*, and *F score*.

Table 2 demonstrates our method significantly outperforms the previous method [7]. The performance gap is more clear when data contains more mitotic or apoptotic(dead) cells. Our method adopts a combination of different phase retardations to detect various stages of cells.

Table 3 summarizes the performance of our algorithm on cell stage classification per frame. For this evaluation, we manually annotated 193 mitotic cells in Seq1, 396 mitotic cells in Seq2, and 596 apoptotic(or dead) cells in Seq3. The precision is not as high as recall since bright halos are often detected as bright cells. Mitosis and apoptosis are temporal events that occur over several frames; thus, the method that does not utilize temporal information and performs classification purely on per-frame features, like ours, obviously has limitation. However, the high recall indicates that our results can be used to provide a way to extract candidates for mitotic or apoptotic cells, for other methods that detect cellular events in a sequence (not per frame) by exploiting temporal contexts, e.g., [10].

4   Conclusion

In this paper, we propose a phase contrast image restoration method based on the dictionary representation of diffraction patterns. The dictionary corresponds to different phase retardations caused by specimens at different cell stages. We formulate a min-ℓ_1 optimization problem to restore the images and propose an iterative algorithm to solve it. Experiments validate that our proposed method outperforms the previous method [7], particularly when cells undergo various stages. High quality restoration can benefit automated cell tracking and cell stage classification.

References

1. Murphy, D.: Phase Contrast Microscopy and Dark-Field Micorscopy. In: Fundamentals of Light Microscopy and Electronic Imaging, pp. 97–116. Wiley (2001)
2. Neumann, B., Held, M., Liebel, U., Erfle, H., Rogers, P., Pepperkok, R., Ellenberg, J.: High-throughput rnai screening by time-lapse imaging of live human cells. Nature Methods 3, 385–390 (2006)
3. Li, K., Miller, E.D., Chen, M., Kanade, T., Weiss, L.E., Campbell, P.G.: Cell population tracking and lineage construction with spatiotemporal context. Medical Image Analysis 12, 546–566 (2008)
4. Kachouie, N.N., Fieguth, P., Jervis, E.: Watershed deconvolution for cell segmentation. In: 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 375–378 (2008)
5. Zimmer, C., Labruyere, E., Meas-Yedid, V., Guillen, N., Olivo-Marin, J.C.: Segmentation and tracking of migrating cells in videomicroscopy with parametric active contours: a tool for cell-based drug testing. IEEE Transactions on Medical Imaging 21, 1212–1221 (2002)
6. Padfield, D., Rittscher, J., Thomas, N., Roysam, B.: Spatio-temporal cell cycle phase analysis using level sets and fast marching methods. Medical Image Analysis 13, 143–155 (2009)
7. Yin, Z., Li, K., Kanade, T., Chen, M.: Understanding the Optics to Aid Microscopy Image Segmentation. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part I. LNCS, vol. 6361, pp. 209–217. Springer, Heidelberg (2010)
8. Needell, D., Vershynin, R.: Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit. IEEE Journal of Selected Topics in Signal Processing 4, 310–316 (2010)
9. Sha, F., Lin, Y., Saul, L.K., Lee, D.D.: Multiplicative updates for nonnegative quadratic programming. Neural Comput. 19, 2004–2031 (2007)
10. Huh, S., Ker, D.F.E., Bise, R., Chen, M., Kanade, T.: Automated mitosis detection of stem cell populations in phasecontrast microscopy images. IEEE Trans. Med. Imaging 30, 586–596 (2011)

Context-Constrained Multiple Instance Learning for Histopathology Image Segmentation

Yan Xu[1,2], Jianwen Zhang[2], Eric I-Chao Chang[2], Maode Lai[4],
and Zhuowen Tu[2,3]

[1] State Key Laboratory of Software Development Environment,
Key Laboratory of Biomechanics and Mechanobiology of Ministry of Education,
Beihang University, China
[2] Microsoft Research Asia, China
[3] Lab of Neuro Imaging, Department of Neurology and Department of Computer
Science, UCLA, USA
[4] Department of Pathology, School of Medicine, Zhejiang University, China
zhuowent@microsoft.com

Abstract. Histopathology image segmentation plays a very important role in cancer diagnosis and therapeutic treatment. Existing supervised approaches for image segmentation require a large amount of high quality manual delineations (on pixels), which is often hard to obtain. In this paper, we propose a new algorithm along the line of weakly supervised learning; we introduce context constraints as a prior for multiple instance learning (ccMIL), which significantly reduces the ambiguity in weak supervision (a 20% gain); our method utilizes image-level labels to learn an integrated model to perform histopathology cancer image segmentation, clustering, and classification. Experimental results on colon histopathology images demonstrate the great advantages of ccMIL.

1 Introduction

High resolution histopathology images provide critical information for cancer diagnosis and analysis [1]. Some clinical tasks for the histopathology image analysis may include [2]: (1) diagnosing the presence of cancer (image classification); (2) segmenting images into cancer and non-cancer cells (image segmentation); (3) clustering the tissue cells into various sub-classes. In this paper, we focus on the segmentation task but our integrated framework essentially is able to perform classification, segmentation, and clustering altogether.

Standard unsupervised image segmentation methods [3] may not work well for the histopathology cancer images due to their complicated patterns. Most of the existing supervised approaches [4] for tissue cell segmentations require detailed manual annotations; this task is not only time-consuming but also intrinsically ambiguous, even for well-trained experts. Recent development in weakly-supervised learning (WSL), more specifically multiple instance learning (MIL) [5], uses coarse-grained labeling to aid automatic exploration of fine-grained information. In MIL, a training set consists of bags (images in our case); each

bag consists of a number of instances (patches in our case); only bag-level labels are given in training; the training algorithm then automatically explores instance-level and bag-level models to best fit the given bag labels. Encouraging results have been observed in medical image classifications [6]; in other medical applications, a multiple instance learning approach was adopted in [7] to detect accurate pulmonary embolism among the candidates; in [8] a CAD system was proposed for polyp detection with the main focus on supervised learning features, which are then used in multiple instance regression; MIL-based histopathology image classification was tackled in [9]. However, none of the above methods were targeted for image segmentation, which is also different from an integrated framework of segmentation, clustering, and classification.

In this paper, we design an MIL-based histopathology image segmentation formulation by proposing a new framework, context-constrained multiple instance learning (ccMIL); it significantly reduces the ambiguity due to the independence assumptions in the standard multiple instance learning. We observe great improvement of ccMIL over competing methods. Our approach also differs from existing formulations in machine learning in the following aspects: latent conditional random fields algorithm [10] deals mostly with compositional components of object models but not for segmentation; MIL on structured data was proposed in [11] but we emphasize the contextual information of instances as a prior here; multiple clustered instance learning (MCIL) [12] adopts the clustering concept into MIL but it takes the assumption of independent instances; a context-based learning/segmentation framework was proposed in [13] but it is a fully supervised approach.

2   Methods

Rich contextual information has important significance for accurate segmentation [13]. ccMIL aims to take into consideration such contextual information to enhance the performance and achieve robustness. An integrated framework, multiple clustered instance learning (MCIL) [12], was recently proposed to perform simultaneous image-level classification, pixel-level segmentation and patch-level clustering. ccMIL inherits some aspects of MCIL but studies the contextual prior in the MIL training stage to reduce the intrinsic ambiguity due to the nature of weak supervision. We observe significant improvement of ccMIL over MCIL in experiments, e.g. over 20% gain. Fig. 1 illustrates the distinction between standard supervised learning, MIL, MCIL and ccMIL.

2.1   Context-Constrained Multiple Instance Learning (ccMIL)

In ccMIL, learning examples are represented by a bag of instances. In our case, a histopathology image is a bag and each patch sampled from an image is an instance. Patches with cancer tissues are treated as positive instances and the ones without cancer tissues are negative. A bag is labeled as positive (cancer image) if the bag contains at least one positive instance.
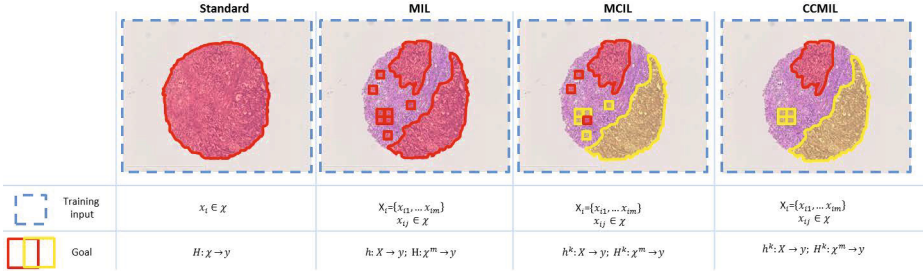
Fig. 1. Distinct learning goals between supervised learning, MIL, MCIL and ccMIL. ccMIL makes an important step over MIL by studying the contextual prior information among the instances to reduce the instance-level ambiguity due to weak-supervision.

We are given a training set $\mathcal{X}$, and x_i is the ith bag in $\mathcal{X}$. Each bag consists of a set of m instances *i.e.* $x_i = \{x_{i1}, \ldots, x_{im}\}$; each x_i is also associated with a label $y_i \in \mathcal{Y} = \{-1, 1\}$. Assume there are K clusters (cancer types), then each instance x_{ij} has a corresponding label $y_{ij}^k \in \mathcal{Y} = \{-1, 1\}$, $k \in \{1, \ldots, K\}$, that denotes whether the instance belongs to the kth cluster. If this instance belongs to one of the K clusters, that is $y_{ij}^k = 1$, then this instance is considered as positive. Note that, this y_{ij}^k is not known during training. A bag is labeled positive if at least one of its instances belongs to one of the K groups:

$$y_i = \max_j \max_k (y_{ij}^k). \tag{1}$$

The goal of ccMIL is to split the positive instance into K groups by learning K instance-level classifiers $h^k(x_{ij}) : \mathcal{X} \to \mathcal{Y}$ for K clusters, using only bag labels y_i, such that $\max_j \max_k h^k(x_{ij}) = y_i$.

We combine AnyBoost[14,15,16] framework, the same as MIL-Boost [15], to solve $h^k(x_{ij})$. First loss function $\mathcal{L}(h)$ (details are given in the next subsection) is introduced to find the optimal weak classifier response $h_t^k : \mathcal{X} \to \mathcal{Y}$ that most reduces the loss on the training data. We train h_t^k by minimizing the training data error weighted by $|w_{ij}^k|$: $h_t^k = argmin_h \sum_{ij} \mathbf{1}(h(x_{ij}^k) \neq y_i^k)|w_{ij}^k|$; while $w_{ij}^k \equiv -\frac{\partial \mathcal{L}(h)}{\partial h_{ij}^k}$. A differentiable *softmax* function $g_l(v_l)$ is given to approximate the max over $\mathbf{v} = \{v_1, \ldots, v_m\}$. It is defined as follows:

$$g_l(v_l) \approx \max_l(v_l) = v^*, \quad \frac{\partial g_l(v_l)}{\partial v_l} \approx \frac{\mathbf{1}(v_i = v^*)}{\sum_l \mathbf{1}(v_l = v^*)}, \quad m = |\mathbf{v}|. \tag{2}$$

There are a number of approximations for g. We choose GM model [16], that is $g_l(v_l) = (\frac{1}{m} \sum_l v_l^r)^{\frac{1}{r}}$, based on the experiment results. In order to optimize the loss function $\mathcal{L}$, we must get p_i, which is defined as the maximum over p_{ij}^k, the probability of an instance x_{ij} belonging to the kth cluster: $p_{ij}^k = \sigma(2h_{ij}^k)$, where $h_{ij}^k = h^k(x_{ij})$. Using the *softmax* g in place of the max, we can get p_i as:

$$p_i = g_j(g_k(p_{ij}^k)) = g_{jk}(p_{ij}^k) = g_{jk}(\sigma(2h_{ij}^k)), \sigma(v) = \frac{1}{1 + \exp(-v)}. \tag{3}$$

So, the weights w_{ij} can be written as $w_{ij}^k = -\frac{\partial \mathcal{L}(h)}{\partial h_{ij}^k} = -\frac{\partial \mathcal{L}(h)}{\partial p_i} \frac{\partial p_i}{\partial p_{ij}^k} \frac{\partial p_{ij}^k}{\partial h_{ij}^k}$. Thus, we can train the weak classifier h_t^k by optimizing weighed error $|w_{ij}^k|$, and finally, get a strong classifier: $\mathbf{h}^k \leftarrow \mathbf{h}^k + \alpha_t{}^k h_t^k$, where α_t weighs the weak learners relative importance.

2.2   Loss Function and Solving Process of ccMIL

The key to ccMIL is a formulation for introducing the contextual information as a prior for MIL.

Now we define two functions $\mathcal{L}_A(h)$ and $\mathcal{L}_B(h)$ as:

$$\mathcal{L}_A(h) = -\sum_{i=1}^{n} w_i(\mathbf{1}(y_i = 1)\log p_i + \mathbf{1}(y_i = -1)\log(1 - p_i)), \quad and \qquad (4)$$

$$\mathcal{L}_B(h) = \sum_{i=1}^{n} w_i \sum_{(j,m)\in E_i} v_{jm} \parallel p_{ij} - p_{im} \parallel^2, \qquad (5)$$

where w_i is the weight of the ith training data (the ith bag). E_i denotes the set of all the neighboring instance pairs in the ith bag. v_{jm} is the weight on a pair of instances (patches) j and m related to the distance (on the image, denoted as d_{jm}) between them. Higher weights are put on those closer pairs. In our experiment, we chose: $v_{jm} = \exp(-d_{jm})$.

Then, we can define loss function as:

$$\mathcal{L}(h) = \mathcal{L}_A(h) + \lambda \mathcal{L}_B(h). \qquad (6)$$

$\mathcal{L}_B(h)$ imposes an effective contextual constraints (in a way smoothness prior) over the instances to remove the ambiguity in training; it encourages the nearby image patches to share similar class types. λ is the weight of the additional item that reflects the importance of relationship between the current instance and its context (neighbors). The overall classification function obtained with the new formulation is thus robust to noise and able to achieve more accurate segmentation results.

According to the new loss function we compute the weight w_{ij}^k as following:

$$w_{ij}^k = -\frac{\partial \mathcal{L}(h)}{\partial h_{ij}^k} = -\frac{\partial \mathcal{L}(h)}{\partial p_i} \frac{\partial p_i}{\partial p_{ij}^k} \frac{\partial p_{ij}^k}{\partial h_{ij}^k}. \qquad (7)$$

$$\frac{\partial \mathcal{L}(h)}{\partial h_{ij}^k} = \frac{\partial \mathcal{L}_A(h)}{\partial h_{ij}^k} + \lambda \frac{\partial \mathcal{L}_B(h)}{\partial h_{ij}^k} = \frac{\partial \mathcal{L}_A(h)}{\partial p_i} \frac{\partial p_i}{\partial p_{ij}^k} \frac{\partial p_{ij}^k}{\partial h_{ij}^k} + \lambda \frac{\partial \mathcal{L}_B(h)}{\partial p_{ij}^k} \frac{\partial p_{ij}^k}{\partial h_{ij}^k}. \qquad (8)$$

$$\frac{\partial \mathcal{L}_A(h)}{\partial p_i} = \begin{cases} -\dfrac{w_i}{p_i} & if \ y = 1; \\[2mm] \dfrac{w_i}{1 - p_i} & if \ y = -1, \end{cases} \qquad \frac{\partial \mathcal{L}_B(h)}{\partial h_{ij}^k} = w_i \sum_{(j,m)\in E_i} 2v_{jm}(p_{ij}^k - p_{im}^k).$$

$$(9)$$

Details of ccMIL are demonstrated in Algorithm 1. K is the number of cancer types, and T is the number of weak classifiers in Boosting.

Algorithm 1. ccMIL

Input: Bags $\{X_1, \ldots, X_n\}, \{y_1, \ldots, y_n\}, K, T$
Output: $\mathbf{h}^1, \ldots, \mathbf{h}^K$
for $t = 1 \rightarrow T$ **do**
   for $k = 1 \rightarrow K$ **do**
      Compute weights $w_{ij}^k = -\frac{\partial \mathcal{L}(h)}{\partial h_{ij}^k} = -(\frac{\partial \mathcal{L}_A(h)}{\partial p_i} \frac{\partial p_i}{\partial p_{ij}^k} \frac{\partial p_{ij}^k}{\partial h_{ij}^k} + \lambda \frac{\partial \mathcal{L}_B(h)}{\partial p_{ij}^k} \frac{\partial p_{ij}^k}{\partial h_{ij}^k})$
      Train weak classifiers h_t^k using weights $|w_{ij}^k|$
      $h_t^k = argmin_h \sum_{ij} \mathbf{1}(h(x_{ij}^k) \neq y_i^k) |w_{ij}^k|$
      Find α_t via line search to minimize $\mathcal{L}(., \mathbf{h}^k, .)$
      $\alpha_t^k = argmin_\alpha \mathcal{L}(., \mathbf{h}^k + \alpha h_t^k, .)$
      Update strong classifiers $\mathbf{h}^k \leftarrow \mathbf{h}^k + \alpha_t^k h_t^k$
   end for
end for

---

3   Experiments

ccMIL is a general approach for common cancer types, including colon, prostate, and breast cancer. Without loss of generality, colon histopathology images are chosen in our experiments to illustrate its effectiveness. We collected the image dataset in Department of Pathology of Zhejiang University in September 2010. The images are obtained from the Nano Zoomer 2.0HT digital slice scanner produced by Hamamatsu Photonics with a magnification factor of 40. In this dataset, 30 non-cancer (NC) images and 53 cancer images are included. The cancer images can be medically divided into four cancer types according to their morphological characteristics. These four cancer types are Moderately or well differentiated tubular adenocarcinoma (MTA), Poorly differentiated tubular adenocarcinoma (PTA), Mucinous adenocarcinoma (MA), and Signet-ring carcinoma (SRC). To ensure the ground truth of the image dataset, images are carefully studied and labeled by experts. Specifically, each image is independently labeled by two pathologists, the third pathologist moderates their discussion until they get an agreement on the result. All images are labeled as cancer images or non-cancer images. For cancer images, cancer tissues are further annotated and corresponding cancer type is identified for the evaluation.

We combine all the images to generate three different subsets: *binary*, *multi*1, and *multi*2. Each subset contains 60 different histopathology images. *binary* contains 30 non-cancer and 30 MTA cancer images. It is used to test the capability of cancer image detection. *multi*1 and *multi*2 mean two or more types of cancer images as well as non-cancer images are contained. They can reveal the ability of pixel-level segmentation. In particular, *multi*1 consists of 30 NC, 15 MTA, 9 PTA and 6 SRC; *multi*2 consists of 30 NC, 13 MTA, 9 PTA and 8 MA. Settings are made as following. First we down-sample the images by 5 times, and then extract 64×64 patches from each image. The parameters in algorithm are set as: $r = 20$, $K = 4$, $T = 200$. r controls sharpness and accuracy in GM model of *softmax* function. The λ used in the loss function is set to 0.01 according to the results of cross validation. We assume the initial distribution is uniform so that
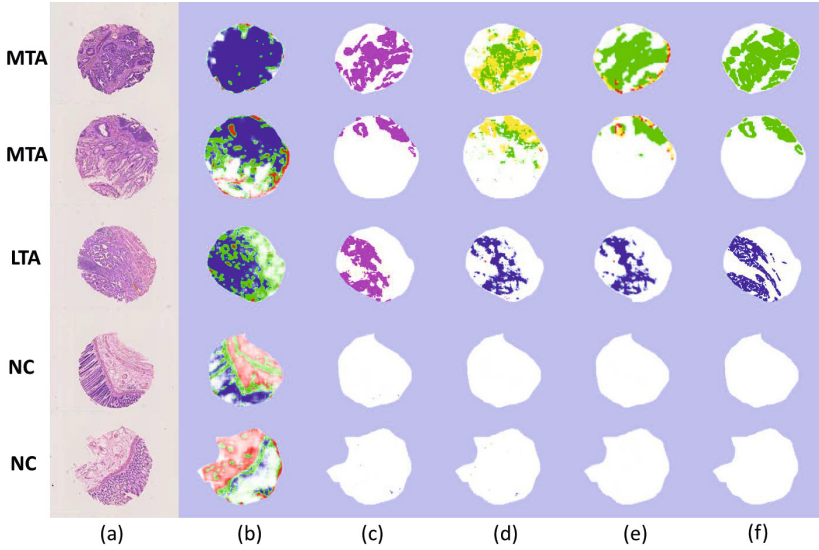
Fig. 2. Image Types: (a): The original images. (b): The instance-level results (pixel-level segmentation and patch-level clustering) for image-level supervision + K-means, (c): pixel-level full supervision, (d): MCIL, (e): ccMIL, (f): The instance-level ground truth labeled by three pathologists. Different colors stand for different types of cancer tissues. Cancer Types: from top to bottom: MTA, MTA, PTA, NC, and NC.

the prior weight w_i for the ith bag is set as the same value. Our method is not focusing on feature design, so generic features for object classification are used here, including $L^*a^*b^*$ Color Histogram, Local Binary Pattern, and SIFT. The weak classifier we use is Gaussian function. Experimental results are reported in a 5-fold cross validation. All the methods in the following experiments are conducted under the same experimental settings.

Pixel-Level Segmentation. We tested subset *multi*2 with different methods to measure pixel-level segmentation. Fig. 2 shows the segmentation results. ccMIL significantly improves results by reducing the intrinsic training ambiguity compared to other weakly supervised methods. For example, ccMIL can correctly recognizes noises and small isolated areas in cancer images and achieve cleaner boundaries, which can be observed from the segmented results of MTA and PTA cancer images in the figure. Moreover, due to the guidance of contextual information, ccMIL reduces the possibility of extracting noises as a positive instance from a non-cancer image and further improves the accuracy of cancer detection.

For the quantitative evaluation of the segmentations, F-measure is used here to evaluate the segmentation. the F-measure values of image-level supervision, MCIL and ccMIL are 0.312, 0.601 and 0.717. ccMIL improves F-measure by 20%, compared with the closet competing method.
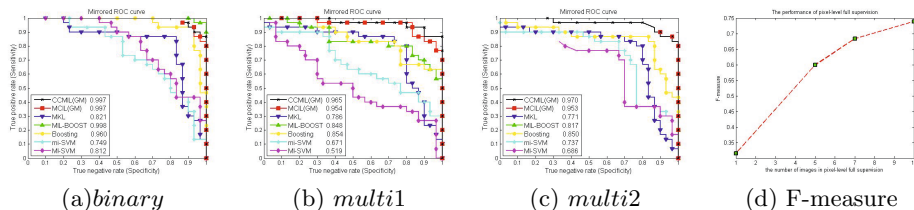
Fig. 3. ROC curves for different learning methods (*binary*, *multi*1 and *multi*2) and (d) the segmentation F-measure of pixel-level full supervision

For comparing with supervised approaches, we implemented two methods: (1) one utilizes supervision in the image level by treating all the pixels in the positive and negative bags as positive and negative instances respectively, and (2) one with the full pixel-level supervision (require laborious labeling work). The advantage of ccMIL over the image-level supervision is proved by both segmented images and F-measure evaluation. As for full pixel-level supervision, in order to compare the performance, we used varying numbers $(1, 5, 7, 10)$ of images of pixel-level fully supervision, and calculated the corresponding values of F-measure. The figure is plotted in Fig. 3.d, from which, it can be concluded that ccMIL is able to achieve comparable results (the value of F-measure is around 0.7).

Patch-Level Clustering. We also obtained the clustering results of the same test data mentioned in pixel-level segmentation and the results are shown in Fig. 2. ccMIL achieves less noisy clustering results than MCIL. Also, it revises the error caused by MCIL, which can be observed from the results obtained from the two MTA images in Fig. 2.

Image-Level Classification. Bag-level classification, that is cancer and non-cancer images classification, is compared in this experiment also. Seven methods, namely MI-SVM, mi-SVM, Boosting, MIL-BOOST, MKL (multiple kernel learning as a widely used image categorization technique), MCIL, and ccMIL, are compared in this experiment with the same features and parameters (we do not put all the references due to the space limit). Fig. 3 shows the receiver operating characteristic (ROC) curves in the three subsets. The results demonstrate the practicality of ccMIL.

4   Conclusion

We have introduced the context constraints to the multiple instance learning framework for segmentation and observe significant improvement (20%) over the closest competing method. In addition, ccMIL is able to perform segmentation, clustering, and classification in a principled framework while achieving comparable results in segmentation with full pixel-level supervision approaches.

References

1. Gurcan, M., Boucheron, L., Can, A., Madabhushi, A., Rajpoot, N., Yener, B.: Histopathological image analysis: A review. IEEE Reviews in Biomedical Engineering 2, 147–171 (2009)
2. Yang, L., Tuzel, O., Meer, P., Foran, D.J.: Automatic Image Analysis of Histopathology Specimens Using Concave Vertex Graph. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 833–841. Springer, Heidelberg (2008)
3. Chan, T.F., Vese, L.A.: Active contours without edges. IEEE Transactions on Image Processing 10(2), 266–277 (2000)
4. Madabhushi, A.: Digital pathology image analysis: opportunities and challenges. Imaging in Medicine 1(1), 7–10 (2009)
5. Maron, O., Lozano-Pérez, T.: A framework for multiple-instance learning. In: NIPS (1997)
6. Liu, Q., Qian, Z., Marvasty, I., Rinehart, S., Voros, S., Metaxas, D.N.: Lesion-Specific Coronary Artery Calcium Quantification for Predicting Cardiac Event with Multiple Instance Support Vector Machines. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part I. LNCS, vol. 6361, pp. 484–492. Springer, Heidelberg (2010)
7. Liang, J., Bi, J.: Computer Aided Detection of Pulmonary Embolism with Tobogganing and Mutiple Instance Classification in CT Pulmonary Angiography. In: Karssemeijer, N., Lelieveldt, B. (eds.) IPMI 2007. LNCS, vol. 4584, pp. 630–641. Springer, Heidelberg (2007)
8. Lu, L., Bi, J., Wolf, M., Salganicoff, M.: Effective 3D object detection and regression using probabilistic segmentation features in CT images. In: CVPR (2011)
9. Dundar, M., Badve, S., Raykar, V., Jain, R., Sertel, O., Gurcan, M.: A multiple instance learning approach toward optimal classification of pathology slides. In: ICPR (2010)
10. Quattoni, A., Wang, S., Morency, L., Collins, M., Darrell, T.: Hidden conditional random fields. IEEE Trans. PAMI 29(10), 1848–1852 (2007)
11. Zhang, D., Liu, Y., Si, L., Zhang, J., Lawrence, R.D.: Multiple instance learning on structured data. In: NIPS (2011)
12. Xu, Y., Zhu, J.-Y., Chang, E., Tu, Z.: Multiple clustered instance learning for histopathology cancer image segmentation, classification and clustering. In: CVPR (2012)
13. Tu, Z., Bai, X.: Auto-context and its application to high-level vision tasks and 3D brain image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 21(10), 1744–1757 (2010)
14. Mason, L., Baxter, J., Bartlett, P., Frean, M.: Boosting algorithms as gradient descent. In: NIPS. MIT Press (2000)
15. Viola, P.A., Platt, J., Zhang, C.: Multiple instance boosting for object detection. In: NIPS (2005)
16. Babenko, B., Dollár, P., Tu, Z., Belongie, S.: Simultaneous learning and alignment: Multi-instance and multi-pose learning. In: Workshop of RealFaces (2008)

Structural-Flow Trajectories for Unravelling 3D Tubular Bundles

Katerina Fragkiadaki[1], Weiyu Zhang[1], Jianbo Shi[1], and Elena Bernardis[2]

[1] Department of Computer and Information Science, GRASP Laboratory,
University of Pennsylvania, Philadelphia, PA 19104
[2] Department of Radiology, University of Pennsylvania, Philadelphia, PA 19104
{katef,zhweiyu,jshi,elber}@seas.upenn.edu

Abstract. We cast segmentation of 3D tubular structures in a bundle as partitioning of *structural-flow trajectories*. Traditional 3D segmentation algorithms aggregate local pixel correlations incrementally along a 3D stack. In contrast, structural-flow trajectories establish *long range* pixel correspondences and their affinities propagate grouping cues across the entire volume simultaneously, from informative to non-informative places. Segmentation by trajectory clustring recovers from persistent ambiguities caused by faint boundaries or low contrast, common in medical images. Trajectories are computed by linking successive registration fields, each one registering pairs of consecutive slices of the 3D stack. We show our method effectively unravels densely packed tubular structures, without any supervision or 3D shape priors, outperforming previous 2D and 3D segmentation algorithms.

Keywords: 3D tubular structures, trajectory clustering, morphological segmentation.

1 Introduction

Automatic segmentation of tubular structures is of vital importance for various fields of medical research. An example of such tubular structures are the organ-pipe-like

a: input 3D stack   **b1:** structural trajectories   **c:** trajectory affinities   **d:** our results 3D tubular segmentation

b2: 2D convexity features

ambiguities

attraction
repulsion
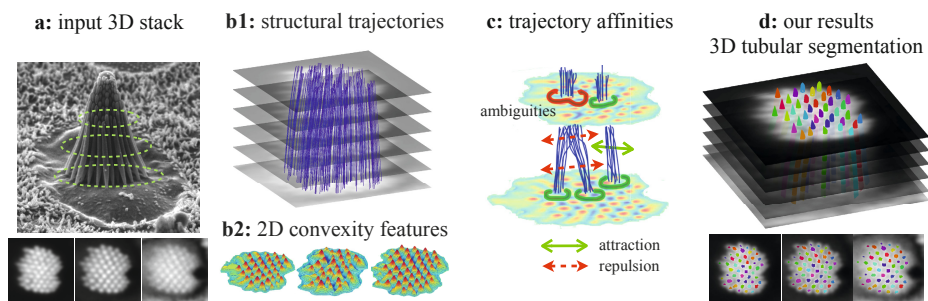
Fig. 1. Method overview. (a) A stack of 2D images of a tubular bundle. Image is courtesy of Medha Pathak and David Corey, Harvard. (b1) Structural-flow trajectories traversing the stack. (b2) 2D convexity cues. (c) Trajectory attractions and repulsions. (d) Resulting 3D segmentation.

stereocilia bundles of the inner ear, depicted in Fig. 1(a). Automatic segmentation of stereocilia in their fluorescent image stacks contributes to medical research on hearing [1].

There are two main lines of work that tackle segmentation of tubular forms: 1) Methods that compute a series of independent 2D segmentations [2–5] and then correspond them along the third dimension [6]. 2) Methods that segment directly in 3D, such as level sets, 3D watershed [4, 3, 7], region growing [8], or methods that employ 3D shape priors, often initialized via some type of user interaction [9–11]. In the former approaches, segmentation and correspondence do not interact with or benefit from each other, hence 2D segmentation mistakes often propagate to erroneous 3D correspondences. In the latter, local correlations along the third dimension are often aggregated in an incremental, feed-forward fashion. Consequently, close configurations between adjacent tubular structures that cause segmentation ambiguity to persist across multiple slices in the 3D image stack are hard to deal with.

Our main insight is that the topology of tubular structures, each with a corresponding one dimensional medial axis and a deforming continuum of 2D cross-sections along the axis direction, allows reliable registration of consecutive cross-sections. A condition for this is the medial axes directions to be non-parallel to the slicing direction. Linking of successive registration fields results in long range pixel correspondences in the 3D volume, which we call *structural-flow trajectories*. We segment densely packed tubular structures by partitioning structural-flow trajectories, as shown in Fig. 1. Trajectory affinities are computed by marginalizing corresponding convexity-driven pixel affinities across trajectory lifespans (Fig. 1(c)). They propagate grouping information along the 3D image stack, from informative to non-informative places and are robust to locally ambiguous grouping cues, often caused by closely attached tubular structures in a bundle. In this way, trajectory partitioning effectively unravels tubular structures automatically (Fig. 1(d)), without 3D shape priors or user interactions.

We test our algorithm on segmenting stereocilia bundles of the inner ear in their fluorescent images. We significantly outperform various baseline segmentation algorithms that do not exploit long range structural information. To the best of our knowledge, we are the first to utilize structural trajectories for capturing long range structural correspondences between pixels at different depths of a 3D volume rather than temporal correspondences between pixels of consecutive frames in a video sequence [12].

2   Long Range Structural Correspondence

Consider two consecutive slices, $I^z(x,y)$ and $I^{z+1}(x,y)$, where $z \in \mathbb{Z}^+$ denotes the slice index from bottom to top of a 3D stack. We define (u,v) to be the deformation field that registers the two slices as the one minimizing intensity and gradient pixel matching scores:

$$\operatorname*{minimize}_{u,v} \quad \begin{aligned} &|I^{z+1}(x+u, y+v) - I^z(x,y)| \\ &+ |\nabla I^{z+1}(x+u, y+v) - \nabla I^z(x,y)| + |\nabla u| + |\nabla v|. \end{aligned} \tag{1}$$

The last two robust penalization terms on gradients of the deformation field u, v encourage smoothness in registration [13]. Such smoothness constraints allow registration to

be reliably computed even in places of ambiguous grouping cues (e.g. faint boundaries), by propagating registration information from reliable (gradient rich) pixel neighbours with peaked unary matching terms. We solve for (u, v) through a coarse to fine estimation scheme with successive linearisation of the intensity and gradient constancy constraints under the assumption of small displacements, as proposed in [13]. Dense slice sampling with respect to deformation along the medial axis of the tubular structures guarantees displacements to be small from slice to slice.

We define a *structure-flow trajectory* to be a sequence of (x, y, z) points:

$$\text{tr}_i = \{(x_i^z, \, y_i^z, \, z), \, z \in \text{Z}_i\}, \tag{2}$$

where Z_i is the set of slice indices in which trajectory tr_i is "alive". Trajectories are dense in space and capture slice-to-slice pixel correspondences, despite illumination changes or density variations of the 2D shapes between slices across the stack. We compute structural trajectories by following per slice registration fields computed from Eq. 2 between pairs of consecutive slices. A forward-backward check determines termination of a trajectory [14]. Thus, structural trajectories can have various lifespans and adapt to the varying lengths of the 3D tubular structures (e.g. stereocilia). We visualize structural trajectories in Fig. 1(b1).

The notion of a pixel trajectory has been traditionally used to describe 2D projections of a single physical point in a video sequence [14]. In our case, the notion of a structural trajectory refers to a series of physical points geometrically related via successive registrations.

3 Constrained Segmentation with Structural Flow Trajectories

In 3D segmentation, local cues are often faint and unreliable. Such ambiguities appear in batches rather than randomly scattered along a 3D stack, since the configuration of 2D cross-sections of the tubular bundle cannot change drastically from one slice to another. We address persistency of cue ambiguity by formulating 3D segmentation as spectral partitioning of structural-flow trajectories. We estimate pixel pairwise relationships at each image slice based on local convexity cues proposed in [5]. Trajectory affinities marginalize corresponding pixel relationships. Thus, grouping cues are propagated from informative to non-informative slices and provide a consistent and well-informed segmentation throughout the whole 3D volume.

3.1 Per Image Grouping Cues

Consider pixel p_i and its neighbourhood $N_d(\text{p}_i)$ of radius r_d, as shown in Fig. 2. We define a peak neighbour p_a of p_i to be a pixel in $N_d(\text{p}_i)$ that can be connected to p_i by a straight line of non-decreasing image intensities, of total intensity difference $\text{S}(\text{p}_i, \text{p}_a) = I(\text{p}_a) - I(\text{p}_i)$. Let $f(\text{p}_i)$ be the weighted average direction from p_i to its peak neighbours:

$$f(\text{p}_i) \propto \sum_{\text{p}_a \in \mathcal{P}(\text{p}_i)} \text{S}(\text{p}_i, \text{p}_a)(\mathbf{p}_a - \mathbf{p}_i), \qquad ||f(\text{p}_i)||_2 = 1, \tag{3}$$
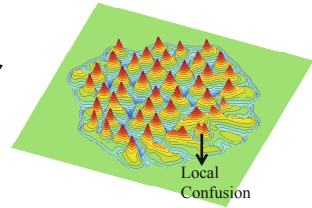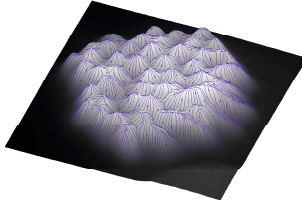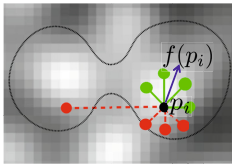
Fig. 2. 2D convexity cues. (a) Estimation of the peak vector $f(\mathrm{p}_i)$. (b) The peak vector field $f(\mathrm{p})$ points at each pixel to the closest highest intensity peak. (c) Degree image D^f. Valleys and peaks correspond to convex and non-convex regions in the original intensity image. Closely attached tubular structures in this slice, cause double valleys in D^f and confuse the corresponding pixel relationships.

where $\mathbf{p}_i$ denotes the 2D pixel coordinate of p_i and $\mathcal{P}(\mathrm{p}_i)$ the set of peak neighbors. We visualize the vector field f in Fig. 2(b).

The inner product of $f(\mathrm{p}_i)$ and $f(\mathrm{p}_a)$ within the neighbourhood $N_d(\mathrm{p}_i)$, measures how much p_a's convexity center agrees with p_i's. We define $\mathrm{D}^f(\mathrm{p}_i)$ to be the sum of such inner products, indicating degree of agreement of a pixel with its surroundings:

$$\mathrm{D}^f(\mathrm{p}_i) = \sum_{\mathrm{p}_a \in N_d(\mathrm{p}_i)} f(\mathrm{p}_i)^\top f(\mathrm{p}_a). \tag{4}$$

We visualize D^f in Fig. 2(c). D^f is rotationally invariant and effectively captures the rough convex shapes of the 2D cross-sections of a tubular structure. Sinks of f (dot centers) are characterized by negative values and sources of f by positive ones. In contrast to morphological charts computed straight from image intensities, D^f is robust to variations of relative intensities of the peaks and valleys in the original image [5].

Given a degree image D^f, we define repulsion $\mathbf{R}_\mathrm{p}(\mathrm{p}_i, \mathrm{p}_j)$ and attraction $\mathbf{A}_\mathrm{p}(\mathrm{p}_i, \mathrm{p}_j)$ between pixels p_i and p_j according to the difference of degrees $\mathrm{D}^f(\mathrm{p}_i)$, $\mathrm{D}^f(\mathrm{p}_j)$ to the minimal degree $m_{ij} = \min_{\mathrm{p}_t \in \mathrm{line}(\mathrm{p}_i, \mathrm{p}_j)} \mathrm{D}^f(\mathrm{p}_t)$ encountered on their connecting line:

$$\begin{aligned} \mathbf{R}_\mathrm{p}(\mathrm{p}_i, \mathrm{p}_j) &= 1 - \exp(-\tfrac{\min(|\mathrm{D}^f(\mathrm{p}_i) - m_{ij}|, |\mathrm{D}^f(\mathrm{p}_j) - m_{ij}|)}{\sigma_r}) \\ \mathbf{A}_\mathrm{p}(\mathrm{p}_i, \mathrm{p}_j) &= (1 - \mathbf{R}_\mathrm{p}(\mathrm{p}_i, \mathrm{p}_j)) \cdot \delta(||\mathrm{p}_i - \mathrm{p}_j||_2 < r_a), \end{aligned} \tag{5}$$

where δ is the delta function. Attractions are short range, acting on pixels within r_a distance. Parameter r_a is chosen as a lower bound of the distance between adjacent structure centers. We set $r_a = 4$ pixels in all our experiments for stereocilia segmentation.

3.2 Trajectory Partitioning

We compute trajectory pairwise relationships by marginalizing pixel relationships across trajectory lifespans. We define repulsion $\mathbf{R}_T(\mathrm{tr}_i, \mathrm{tr}_j)$ between trajectories tr_i and tr_j

to be the maximum of corresponding pixel repulsions and attraction $\mathbf{A}_T(\mathrm{tr}_i, \mathrm{tr}_j)$ to be the minimum of corresponding pixel attractions as follows:

$$
\begin{aligned}
\mathbf{R}_T(\mathrm{tr}_i, \mathrm{tr}_j) &= \max_{z \in Z_i \cap Z_j} \mathbf{R}_{\mathrm{p}}(\mathrm{p}_i^z, \mathrm{p}_j^z) \cdot \delta(|Z_i \cap Z_j| > 0) \\
\mathbf{A}_T(\mathrm{tr}_i, \mathrm{tr}_j) &= \min_{z \in Z_i \cap Z_j} \mathbf{A}_{\mathrm{p}}(\mathrm{p}_i^z, \mathrm{p}_j^z) \cdot \delta(|Z_i \cap Z_j| > 0),
\end{aligned}
\tag{6}
$$

where superscript z indicates the slice index of a pixel. The above cue marginalization reflects the nature of tubular structure bundles: in some slices, tubular structures attached to each other confuse corresponding degree fields as shown in Fig. 2(c), causing leakage in segmentation. On the contrary, over-segmentation of 2D cross sections is highly *unlikely* under our convexity cues. As such, trajectory affinities essentially try to detect the informative slice where attached structures separate.

We classify trajectories as foreground or background by thresholding their average degrees $\mathrm{D}^f(\mathrm{tr}_i) = \mathrm{mean}_{z \in Z_i} \mathrm{D}^f(\mathrm{p}_i^z)$ at 0. Let $X \in \{0,1\}^{|\mathcal{T}| \times K}$ be the matrix whose columns are the indicator vectors of K clusters. We cluster foreground trajectories by maximizing intra-cluster attractions $\mathbf{A}_T$ and inter-cluster repulsions $\mathbf{R}_T$ [15]:

$$
\text{maximize} \qquad \varepsilon(X) = \sum_{k=1}^{K} \frac{X_k^\top (\mathbf{A}_T - \mathbf{R}_T + \mathbf{D}_R) X_k}{X_k^\top (\mathbf{D}_A + \mathbf{D}_R) X_k}
\tag{7}
$$

$$
\text{subject to} \qquad X\mathbf{1}_{|\mathcal{T}|} = \mathbf{1}_{|\mathcal{T}|},
$$

where $\mathbf{D}_A = \mathrm{Diag}(\mathbf{A}_T \mathbf{1}_{|\mathcal{T}|})$, $\mathbf{D}_R = \mathrm{Diag}(\mathbf{R}_T \mathbf{1}_{|\mathcal{T}|})$ are degree matrices and $\mathbf{1}_{|\mathcal{T}|}$ is the $|\mathcal{T}| \times 1$ vector of 1. We choose K to be a rough upper-bound of the total number of tubular structures present in the stack, estimated from the per frame degree fields. We obtain the near-global optimal continuous solution of Eq. 7 from the top K generalized eigenvectors of $(\mathbf{A}_T - \mathbf{R}_T + \mathbf{D}_R, \mathbf{D}_A + \mathbf{D}_R)$. We discretize the eigenvectors by rotation [16] and obtain K clusters. We repeatedly merge clusters with no repulsion between their trajectories. Structure bifurcation can be accommodated by a hierarchical segmentation scheme, where cluster merging probabilities depend on ratios of cluster attractions versus repulsions. We summarize our method in Algorithm 1.

Algorithm 1. Unraveling Tubular Structures

1: Let $\{I^z,\ z = 1 \cdots T\}$ denote an ordered sequence of T images in a 3D stack.
2: **for all** $I^z, z = 1 \cdots T$ **do**
3:     Compute peak vector field $f^z(\mathrm{p}_i)$ and degree field $\mathrm{D}_z^f(\mathrm{p}_i)$ using Eq. 3 and Eq. 4.
4:     Compute pixel attractions $\mathbf{A}_{\mathrm{p}}$ and repulsions $\mathbf{R}_{\mathrm{p}}$ using Eq. 5.
5: **end for**
6: Compute structural trajectories $\mathrm{tr}_i, i = 1 \cdots |\mathcal{T}|$ using method of [14].
7: Compute trajectory degrees $\mathrm{D}^f(\mathrm{tr}_i),\ i = 1 \cdots |\mathcal{T}|$.
8: Classify trajectories as foreground $\mathcal{T}^F = \{\mathrm{tr}_i | \mathrm{D}^f(\mathrm{tr}_i) > 0\}$.
9: Compute foreground trajectory attractions $\mathbf{A}_T$ and repulsions $\mathbf{R}_T$ using Eq. 6.
10: Compute the top K generalised eigenvectors V of $(\mathbf{A}_T - \mathbf{R}_T + \mathbf{D}_R, \mathbf{D}_A + \mathbf{D}_R)$.
11: Discretize eigenvectors V by rotation [16] to obtain K trajectory clusters $G_i, i = 1 \cdots K$.
12: **while** $\exists\, G_i, G_j, \mathbf{R}_T(G_i, G_j) = 0$ **do**
13:     Merge G_i, G_j
14: **end while**

4   Experiments

We test our method on segmenting stereocilia of the hair cells in the inner ear from their fluorescent image stacks. Each stack is 7 to 20 slices long and contains 50 to 70 stereocilia. 3D ground-truth stereocilia centers are marked in each image stack. Ground-truth samples are illustrated in the first column of Fig. 3. We compare our method against three baseline approaches: 1) 3D k-means on pixel intensities. Number of centers k is chosen to achieve best performance. The resulting clusters are pruned based on their size and aspect ratio. 2) 3D watershed (MATLAB built-in implementation) 3) Dot finding [5] using code provided by the authors. Given the 2D output dots of [5], we produce the 3D segmentation by linking segmented dots between consecutive slices via Hungarian matching. We evaluate performance with the following metrics:
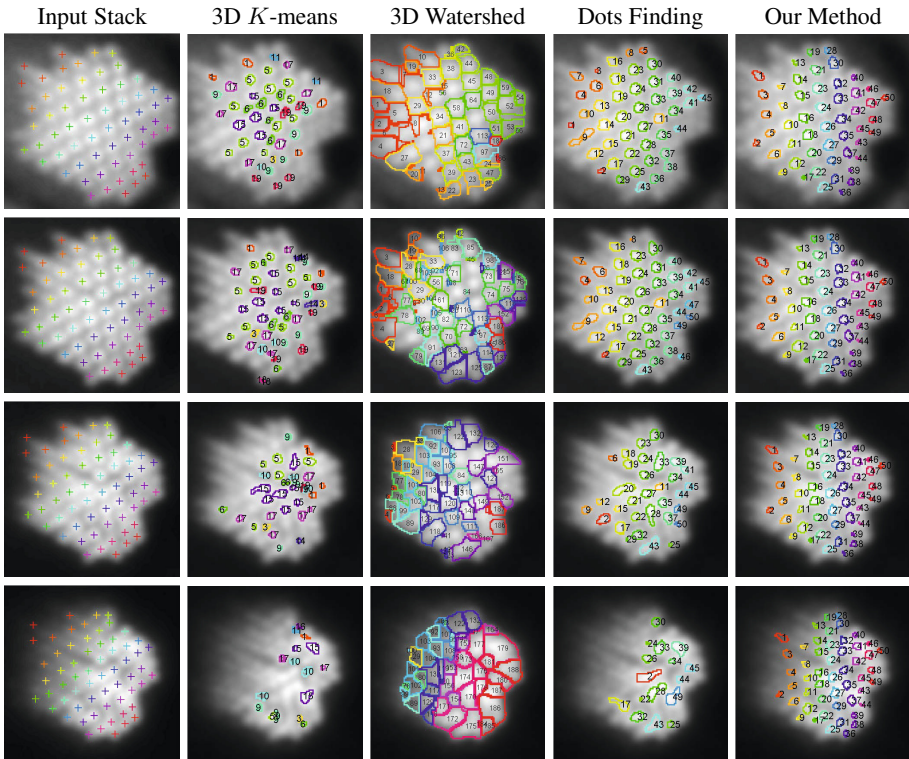
Fig. 3. Segmenting a stereocilia stack (best viewed in color). First column shows 4 (out of 22) images of a stereocilia stack with corresponding 3D ground-truth tubular structure centers. Depth decreases from top to bottom. Column 2-5 show 3D segmentation using 3D k-means, 3D watershed, dot finding [5] and our method respectively. Numbers and colours indicate tubular structure identities. In 3D watershed, tubular structures leak across faint boundaries and break arbitrarily between slices. In dot finding, notice the leaking segments of numbers 9, 20, 38, 43, etc. Our method provides consistent 3D segmentations, correcting leakages and miss-detections.
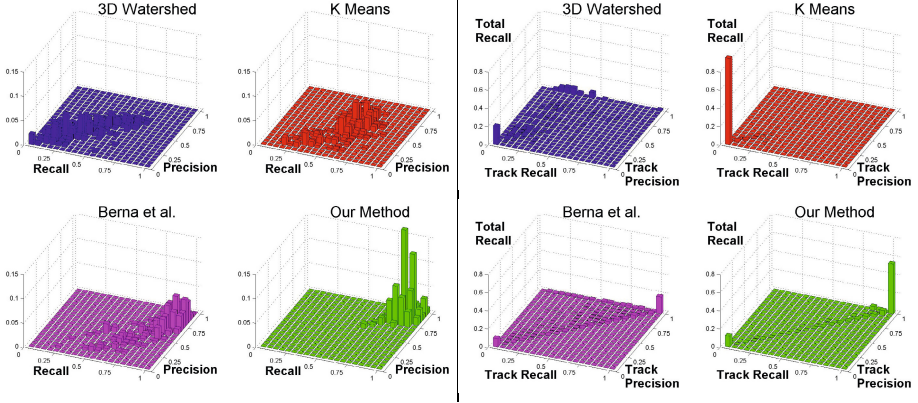
Fig. 4. *Left:* Precision-recall for 2D slice segmentation. We histogram (prec_z, rec_z) values for all slices z in our stacks. *Right:* Precision-recall for 3D tubular structure identification. We histogram (track-rec$_i$, track-prec$_i$) of all labelled tubular structures in our stacks. Best performance is achieved when the histogram is concentrated at the **right top corner** (precision=1, recall=1). Our method (in green) has significantly higher precision and recall in both tasks.

1) **Goodness of 2D segmentation.** For each slice z, given m_z ground-truth structure centers and n_z segment centers hypotheses, let d_{ij}^z be the Euclidean distance between structure center i and segment center j in slice z. We use the following measures:

$$\mathrm{prec}_z = \frac{\#\{j\colon \min_{i=1}^{m_z} d_{ij}^z \leq \tau\}}{n_z}, \qquad \mathrm{rec}_z = \frac{\#\{j\colon \min_{i=1}^{m_z} d_{ij}^z \leq \tau\}}{m_z}. \tag{8}$$

We visualize the histogram of (prec_z, rec_z) over all slices in Fig. 4 *left*. Same evaluation metrics were used in [5].

2) **Goodness of 3D identification.** Given m 3D ground-truth tubular structures and n 3D tubular structures hypotheses, let ℓ_i^g denote the length of ground-truth structure i and ℓ_j^d denote the length of segment structure hypothesis j. We use the following measures:

$$\text{track-rec}_i = \max_{j=1}^n \frac{\#\{z\colon d_{ij}^z \leq \tau\}}{\ell_i^g}, \qquad \text{track-prec}_i = \max_{j=1}^n \frac{\#\{z\colon d_{ij}^z \leq \tau\}}{\ell_j^d}. \tag{9}$$

Tracking precision and recall together quantify how consistently the 3D segmentation hypotheses capture the 3D ground-truth structures [17]. We visualize the histogram of (track-rec$_i$,track-prec$_i$) over all labelled tubular structures in Fig. 4 *right*. We set $\tau = 3$.

Our method outperforms all baseline approaches. Low contrast and faint boundaries make stereocilia segmentation challenging. Our gain in performance comes from corrections of leakages and miss-detections by propagating separations or detections from informative to ambiguous places in the 3D volume, as shown in Fig. 4. Miss-detections in our method are often due to localization errors: segment hypotheses centers are a bit more than 3 pixels away from the corresponding ground-truth. A local gradient descent for discovering the intensity peak in the local neighbourhood could alleviate from such localization mistakes. We did not add this step to keep the method clean.

5   Conclusion

We presented an algorithm for unravelling 3D tubular structures in a tight bundle by propagating grouping information across multiple cross-sections of their 3D volume simultaneously via spectral partitioning of structural-flow trajectories. Our qualitative and quantitative results show our method effectively integrates local grouping cues for accurate segmentation and identification of densely packed structures, outperforming 3D and 2D baseline segmentation algorithms. We are currently exploring ways of applying our algorithm to 4D cell tracking, where both temporal and structural correspondences would mediate cues for segmenting spatio-temporal cell structures.

References

1. Ciuman, R.R.: Auditory and vestibular hair cell stereocilia: relationship between functionality and inner ear disease. Journal of Laryngology Otology 125, 991–1003 (2011)
2. Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 22, 888–905 (2000)
3. Beucher, S.: Mathmatique, C.D.M.: The watershed transformation applied to image segmentation. Scanning Microscopy International 6, 1–26 (1991)
4. Meyer, F.: Morphological segmentation revisited. Space, Structure and Randomness 183, 315–347 (2005)
5. Bernardis, E., Yu, S.X.: Finding dots: Segmentation as poping out regions from boundaries. In: CVPR (2010)
6. Bernardis, E., Yu, S.X.: Structural correspondence as a contour grouping problem. In: Proceedings of IEEE Workshop on MMBIA (2010)
7. Wahlby, C., Sintorn, I.M., Erlandsson, F., Borgefors, G., Bengtsson, E.: Combining intensity, edge and shape information for 2D and 3D segmentation of cell nuclei in tissue sections. Journal of Microscopy 215, 67–76 (2004)
8. Haralick, R., Shapiro, L.: Image segmentation techniques. Computer Vision, Graphics and Image Processing 29(1), 100–132 (1985)
9. 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: Taylor, C.J., Noble, J.A. (eds.) IPMI 2003. LNCS, vol. 2732, pp. 136–147. Springer, Heidelberg (2003)
10. Lorigo, L.M., Grimson, W.E.L., Faugeras, O., Keriven, R., Kikinis, R., Nabavi, A., Westin, C.F.: Codimension - two geodesic active contours for the segmentation of tubular structures. In: CVPR (2000)
11. Dorin, H.T., Tek, H., Comaniciu, D., Williams, J.P.: Vessel detection by mean shift based ray propagation. In: Proceedings of IEEE Workshop on MMBIA (2001)
12. Fragkiadaki, K., Shi, J.: Exploiting motion and topology for segmenting and tracking under entanglement. In: CVPR (2011)
13. Brox, T., Bruhn, A., Papenberg, N., Weickert, J.: High Accuracy Optical Flow Estimation Based on a Theory for Warping. In: Pajdla, T., Matas, J. (eds.) ECCV 2004. LNCS, vol. 3024, pp. 25–36. Springer, Heidelberg (2004)
14. Sundaram, N., Brox, T., Keutzer, K.: Dense Point Trajectories by GPU-Accelerated Large Displacement Optical Flow. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part I. LNCS, vol. 6311, pp. 438–451. Springer, Heidelberg (2010)
15. Yu, S.X., Shi, J.: Understanding popout through repulsion. In: CVPR (2001)
16. Yu, S.X., Shi, J.: Multiclass spectral clustering. In: ICCV (2003)
17. Smith, K., Gatica-perez, D., Odobez, J.-M., Ba, S.: Evaluating multi-object tracking. In: Workshop on Empirical Evaluation Methods in Computer Vision (2005)

Online Blind Calibration of Non-uniform Photodetectors: Application to Endomicroscopy

Nicolas Savoire, Barbara André, and Tom Vercauteren

Mauna Kea Technologies, Paris

Abstract. We present an original method for the online blind calibration of non-uniform photodetectors. The disparity of the detectors may arise from both irregular spatial arrangement and distinct slowly time-varying photometric transfer functions. As natural images are mostly continuous, the signal *collected* by neighboring detectors is strongly correlated over time. The core idea of our method is to translate the calibration problem into relative pairwise calibrations between neighboring detectors followed by the regularized inversion of a system akin to gradient-based surface recovery. From our blind calibration procedure, we design an *online* blind calibration pipeline compatible with clinical practice. *Online blind* calibration is proved to be statistically better than standard *offline* calibration for reconstructing endomicroscopy sequences.

1 The Need for Online Blind Calibration in Endomicroscopy

In vivo endomicroscopy consists of imaging the tissue at microscopic level, by inserting, through a standard endoscope, a probe made of tens of thousands of optical fibers. A laser scanning unit sends along each fiber an excitation light that is locally absorbed by fluorophores in the tissue and emitted back at a longer wavelength along the same fiber to a photodetector. Raw images are produced at a rate of 9 to 18 frames per second. The standard endomicroscopy image reconstruction scheme proposed by Le Goualher et al. [1] is composed of three steps: estimation of the signal collected by each fiber, compensation of calibration coefficients, and interpolation. Alternative less physically motivated methods [2, 3] use image filtering to remove the fiber honeycomb pattern at the cost of introducing some blur. As described in [1], each fiber i of the endomicroscopy probe acts as a mono-pixel photodetector with an affine transfer function:

$$v_i(t) = \alpha_i u_i(t) + \beta_i \tag{1}$$

where α_i and β_i are respectively the gain and offset of fiber i, $u_i(t)$ is the concentration of fluorophore *seen* at time t by fiber i, and $v_i(t)$ is the signal collected at time t by fiber detector i.

For the estimation of the calibration coefficients, Le Goualher et al. [1] proposed a *offline* non-blind calibration method, assuming constant gain and offset for each fiber. This method compensates for affine fiber transfer functions by

previously acquiring two reference images, one in the air and one in a fluo-
rophore solution, and then deducing from these two images a static offset and a
static gain for each fiber. However, fiber coefficients are both slowly time-varying
and medium-dependent. Indeed, due to photo-bleaching, autofluorescence signal
slowly decreases with illumination time during the course of the procedure un-
less a long pre-illumination has been performed. In addition, fiber gains also
depend on refractive indices and dispersive properties of the observed medium.
For these reasons, image quality after the *offline* calibration may be affected by
the presence of a *veil* on the image, i.e. a static noise pattern, which can be seen
in all three images on the top of Fig. 3. Blindly recovering the parameters of non-
uniform photodetectors is an ill-posed problem that has been scarcely addressed
in the literature. In the field of astronomy, Kuhn et al. [4] proposed a blind cali-
bration method that requires shifted images of the exact same scene, which can
typically not be acquired on *in vivo* samples. The closest work is found in the
field of computer vision. Weiss [5] presented an algorithm for blindly recovering
intrinsic images from a sequence of natural scenes with different illumination, but
only considered detectors on regular grids and with a single gain coefficient. To
the best of our knowledge, this paper proposes the first online blind calibration
of non-uniform mono-pixel photodetectors.

Our first objective is to design a blind method for the calibration of pho-
todetectors having both irregular spatial arrangement and distinct photometric
transfer functions. We then show how our blind calibration can be applied in an
online manner to handle slowly time-varying transfer functions.

2   Solving an Inverse Problem on Temporal Windows

Let us first focus on static blind calibration. The core observation behind our
method is that, when looking at physical objects, neighbor fibers nearly *see* the
same *object* signal over time, so the *collected* signal from two neighbor fibers is
strongly temporally correlated. This property arises from the mostly continuous
nature of natural images, where edges form a set of zero measure and where
the noise can be measured for example using total variation as introduced by
Rudin et al. [6]. To translate this into mathematical terms, we capture the spatial
relationship of the irregularly arranged photodetectors with a Delaunay trian-
gulation applied to fiber locations, and express (1) in terms of a global function
$U_t(\boldsymbol{p})$ which represents the concentration of fluorophore in the observed medium
at position $\boldsymbol{p}$ and time t. Let $G = (V, E)$ the undirected graph associated with
the triangulation. In our endomicroscopy probes the fiber pattern is pseudo
hexagonal leading to $|E| \simeq 3|V|$. In (1), we substitute for $u_i(t)$ the expression
$U_t(\boldsymbol{\phi}(t) + \boldsymbol{p_i})$, where $\boldsymbol{p_i}$ is the spatial position of fiber i relative to the center of
probe distal end, and $\boldsymbol{\phi}(t)$ is the spatial position of the probe center at time t.
We then have:

$$v_j(t) = \frac{\alpha_j}{\alpha_i}(v_i(t) - \beta_i) + \beta_j + \epsilon_{ij}(t) \tag{2}$$

where $\epsilon_{ij}(t) = (\boldsymbol{p_j} - \boldsymbol{p_i}) \cdot \nabla U_t(\boldsymbol{p_i} + \boldsymbol{\phi}(t)) + o(\|\boldsymbol{p_j} - \boldsymbol{p_i}\|)$ is a noise term whose
distribution is sparse, with the natural assumption, similar to [6], that spatial

variations of the concentration $U_t(.)$ are almost always smooth. The signals collected by two neighboring fibers can thus be related by the affine relationship $v_j(t) \simeq a_{ij}v_i(t) + b_{ij}$, from which we derive two systems:

$$\{a_{ij} = \frac{\alpha_j}{\alpha_i}\}_{(i,j)\in E} \tag{3}$$

$$\{b_{ij} = \beta_j - \frac{\alpha_j}{\alpha_i}\beta_i\}_{(i,j)\in E} \tag{4}$$

Interestingly, the affine coefficients a_{ij} and b_{ij} may conceptually be seen as the *gradients* of the fiber gain map and of the fiber offset map, respectively, except that the fibers are distributed on a non regular grid. Thus, the problem of recovering the calibration coefficients from the relative pairwise calibrations is akin to recovering a surface from a gradient field [7,8].

Our blind calibration method is designed in two steps. The first step performs linear regressions on a buffer of collected images to estimate the relative calibration coefficients (a_{ij}, b_{ij}) for all $(i, j) \in E$. Once all regressions have been computed, the second step consists of solving the gain system (3) and the offset system (4), in order to recover the gain α_i and the offset β_i of each fiber i.

2.1   Pairwise Relative Calibrations of Photodetectors

The first step of our blind calibration method consists of estimating relative calibration coefficients a_{ij} and b_{ij} between two observed signals $v_i(t)$ and $v_j(t)$ that are collected by neighboring fibers (i, j) and sampled at times $t \in \{1, .., m\}$ without loss of generality. From equation (2), we have: $v_j(t) = a_{ij}v_i(t) + b_{ij} + \epsilon_{ij}(t)$. Noticing that there is measurement error in both variables, an orthogonal linear regression [9] is more appropriate than ordinary least squares regression to estimate a_{ij} and b_{ij}. In order to account for outliers resulting from the non-normality of $\epsilon_{ij}(t)$, the most common example being when an edge occurs between two neighbor fibers, we perform a robust orthogonal regression. For this purpose, we use M-estimators because they are deterministic and computationally competitive compared to other methods. We choose to perform an iteratively reweighted least squares fitting algorithm, which is at iteration p:

$$\hat{\boldsymbol{\theta}}_{ij}^{(p)} = \underset{\{\boldsymbol{\theta}_{ij}|\theta_{1_{ij}}^2+\theta_{2_{ij}}^2=1\}}{\arg\min} \sum_{t=1}^{m} z(r_{ij,t}^{(p-1)})(\theta_{1_{ij}}v_i(t) + \theta_{2_{ij}}v_j(t) + \theta_{3_{ij}})^2 \tag{5}$$

where $\boldsymbol{\theta}$ represents the regression coefficients, $z(x) = \rho'(x)/x$ is the weight function and $r_{ij,t}^{(p)} = \theta_{1_{ij}}^{(p)}v_i(t) + \theta_{2_{ij}}^{(p)}v_j(t) + \theta_{3_{ij}}^{(p)}$ is the orthogonal residual at iteration p associated with a sample at time t. By choosing the Tukey's biweight function for $\rho(.)$, we ensure that the residuals larger than a cutoff value c are eliminated. We set $c = 4.6851\hat{\sigma}$, where $\hat{\sigma}$ is the median absolute deviation of residuals. Our robust algorithm initialization sets initial slope to s_j/s_i and initial intercept to $\mu_j - s_j\mu_i/s_i$, where μ_i and s_i respectively denote the median and interquartile range of $\{v_i(t)\}_{1\leq t\leq m}$.

As the probe may contain *dead* fibers that do not transmit any light, any regression between these fibers and their neighbors is sterile. To overcome this issue, we decide to detect and reject regressions which are bad fits, based on a robust goodness-of-fit measure g_{ij} for $(i,j) \in E$. We define g_{ij} as the *weighted* squared Pearson correlation coefficient between v_i and v_j: $g_{ij} = (\sum_{t=1}^{m} z_{ij,t}(v_i(t) - \overline{v_i})(v_j(t) - \overline{v_j}))^2 / (\sum_{t=1}^{m} z_{ij,t}(v_i(t) - \overline{v_i})^2 \sum_{t=1}^{m} z_{ij,t}(v_j(t) - \overline{v_j})^2)$, where $\{z_{ij,t}\}_{1 \leq t \leq m}$ are the final weights of (5) after convergence, and $\overline{v_i} = (\sum_{t=1}^{m} z_{ij,t} v_i(t))/(\sum_{t=1}^{m} z_{ij,t})$. These $\{g_{ij}\}_{(i,j) \in E}$ will be used in the second step of our blind calibration method, focusing on system inversion.

2.2   From Relative Calibrations to Global Calibration

Gain Estimation. The gain system (3) uses the results of an linear orthogonal regression. The symmetry of orthogonal regression implies that noise on the estimates $\hat{a}_{ij}$ is better modeled as multiplicative. Therefore, (3) can be advantageously transposed to the log domain: $\{\alpha_j^* - \alpha_i^* = \hat{a}_{ij}^*\}_{(i,j) \in E}$ with $\alpha_i^* = \log \alpha_i$ and $\hat{a}_{ij}^* = \log \hat{a}_{ij}$. This can be rewritten as $M\boldsymbol{\alpha}^* = \hat{\boldsymbol{a}}^*$, where M is the $|E| \times |V|$ matrix such as $M_{lp} = \delta_{p,j_l} - \delta_{p,i_l}$, with $E = \{(i_l, j_l), 1 \leq l \leq |E|\}$. In order to cope with noise and *dead* fibers, we introduce the weight function $w(x) = (1 + e^{-(x-g_0)})^{-1}\delta(x \geq g_0)$, where δ is the Kronecker operator. This function puts more weight on the relations of (3) associated with higher goodness-of-fit values, and ignores those associated with values below an arbitrary cutoff value $g_0 = 0.6$. To cope with the non-uniqueness of the solution and regularize the system, we add the a priori that the gains are close to 1. We use the $|E| \times |E|$ diagonal matrix W defined by $W_{ll} = w(g_{i_l j_l})$ to obtain the weighted system:

$$\hat{\boldsymbol{\alpha}}^* = \underset{\boldsymbol{\alpha}^*}{\arg\min} \|W(M\boldsymbol{\alpha}^* - \hat{\boldsymbol{a}}^*)\|^2 + \lambda\|\boldsymbol{\alpha}^*\|^2 \tag{6}$$

The overdetermined system (6) is solved using a conjugate gradient method (LSQR) which has a good numerical stability for ill-conditioned systems. Finally, we deduce the estimated gains $\hat{\boldsymbol{\alpha}} = \exp \hat{\boldsymbol{\alpha}}^*$.

Offset Estimation. In order to uncouple offset estimation from gain estimation, we approximate α_j/α_i by $\hat{a}_{ij}$ in the offset system (4). As (4) uses the estimates $\hat{b}_{ij}$ resulting from the symmetric orthogonal regression, we symmetrize the offset system by introducing a normalization factor $\gamma_{ij} = (1 + \hat{a}_{ij}^2)^{0.5}$: $\beta_j/\gamma_{ij} - (\hat{a}_{ij}/\gamma_{ij})\beta_i = \hat{b}_{ij}/\gamma_{ij}$. In order to regularize the system, we make the assumption $\mathcal{A}$ that fiber background $\boldsymbol{\beta}(t)$ at time $t > t_0$ can be approximated by $q\boldsymbol{\beta}_0$ where $\boldsymbol{\beta}_0$ is an initial offline background estimation and q is an unknown global factor only depending on t [10]. Injecting $\beta_i = q\beta_{0_i}$ in (4) gives: $q\beta_{0_j} - \hat{a}_{ij}q\beta_{0_i} \simeq b_{ij}$. A robust estimate of q is then: $\hat{q} = \text{median}_{(i,j) \in E}\{\hat{b}_{ij}/(\beta_{0_j} - \hat{a}_{ij}\beta_{0_i})\}$. From this estimate $\hat{q}$, we obtain an approximation $\hat{q}\boldsymbol{\beta}_0$ of $\boldsymbol{\beta}$ and use this value to regularize the system (4). Weighting the system as described in 2.2, offset estimation is rewritten as:

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\arg\min} \|W(A\boldsymbol{\beta} - \frac{1}{\gamma}\hat{\boldsymbol{b}})\|^2 + \lambda\|\boldsymbol{\beta} - \hat{q}\boldsymbol{\beta}_0\|^2 \tag{7}$$

where A is the $|E| \times |V|$ matrix such as $A_{lp} = (\delta_{p,j_l} - \delta_{p,i_l}\hat{a}_{i_l j_l})/\gamma_{i_l j_l}$. Being similar to system (6), the system (7) is solved using LSQR.

2.3 Online Blind Calibration Pipeline

Our second objective is to integrate the blind calibration method described in Section 2.2 into an online calibration pipeline, illustrated in Fig. 2. For this purpose, we leverage the fact that the temporal variation of fiber transfer functions are quite slow compared to the temporal variation of fluorescence signal which is due to either motion within the tissue, or motion of the probe along the tissue during endomicroscopy procedure. We thus decide to apply our blind calibration method on temporal windows where transfer functions do not significantly change, and to perform successive calibration updates as soon as possible. At startup of the imaging system, the probe is in the air and laser illumination allows to record the non-fluorescent background signal $\boldsymbol{\beta}_0$. Standard *offline* calibration can be used as initial calibration. When the physician starts the acquisition at time T_0, successive frames are accumulated in a buffer until there is sufficient data to make system inversion possible. Our intuition is that the more moving fluorescent signal there is, the better linear regression results are, and so the more chance there is to solve the system. We consider that there is sufficient data if the following criterion $\mathcal{C}$ is satisfied: $|\{(i,j) \in E, g_{ij} \geq g_0\}| \geq 0.9|E|$. Regressions are performed until time T_1 when $\mathcal{C}$ is satisfied, then we perform system inversion and proceed to the first calibration update. At k^{th} calibration update, buffer is flushed at $t = T_k + 1$ and new frames are accumulated in the buffer from $t = T_k + 1$ to $t = T_{k+1}$ when $(k+1)^{\text{th}}$ calibration is possible. From the gains $\hat{\boldsymbol{\alpha}}^{T_k}$ and the offsets $\hat{\boldsymbol{\beta}}^{T_k}$ estimated at time T_k, we construct the new calibrated signal as follows: $\forall t \in]T_k, T_{k+1}], \hat{u}_i(t) = (v_i(t) - \hat{\beta}_i^{T_k})/\hat{\alpha}_i^{T_k}$. This provides an estimate of the fluorophore concentration $u_i(t)$ seen by fiber i at time t. Finally, interpolation of $\hat{\boldsymbol{u}}$ is performed.

3   Evaluation and Results

Before method evaluation, we aim at validating the assumption $\mathcal{A} : \boldsymbol{\beta} = q\boldsymbol{\beta}_0$. For this purpose, we tested 4 probes having different numbers of fibers and different optical properties. Each probe was used to acquire every minute a 10-frame average image in the air. Laser illumination was continuously on during the first 14 minutes in order to measure autofluorescence decrease. After 14 minutes, to highlight the phenomenon of autofluorescence recovery, illumination was off except during acquisition of the images. We then calculated the Pearson correlation between the background signal of all fibers at time t and those at time $t_0 = 0$. The results presented in Fig. 1 on the right, reveal that, for all 4 probe models the Pearson correlation coefficient is quite high, being superior to 0.98, therefore validating $\mathcal{A}$.

Our database used for method evaluation is composed of 89 endomicroscopy sequences acquired during clinical procedures in 6 different medical centers. Each
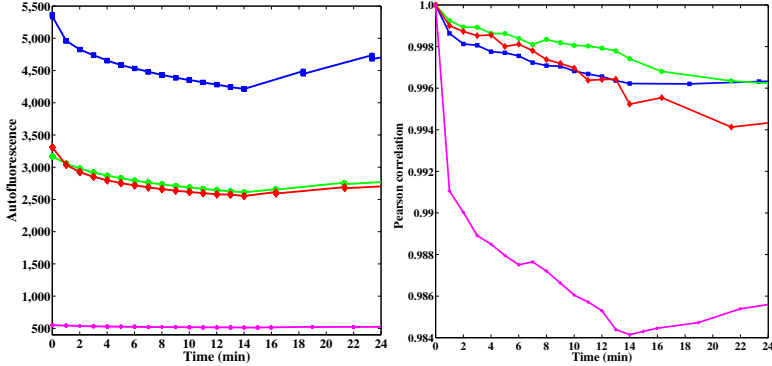
Fig. 1. Left: Autofluorescence decrease ($t \leq 14min$) and recovery ($t > 14min$), measured for 4 different probe models. Right: Corresponding Pearson correlation values between fiber signal at time t and at time t_0.

of these 89 sequences is complete in the sense that every single image from startup to shutdown was recorded. We propose to evaluate first the ability of our calibration method to recover the model parameters, then the impact of the calibration on image reconstruction. Since it is impossible to get an uniform medium with the same physical properties as a given observed living tissue, validation cannot rely on flat-field images.

Our online pipeline implies that calibration applied on *current* data is only computed from *past* data. This ensures unbiased performance evaluation on reconstructed data. Given a temporal window $]T_k, T_{k+1}]$ where the calibration is static, we define a static calibration quality CQ_k which measures the ability to align the calibrated gains of two neighboring fibers from the model parameters estimated on $]T_{k-1}, T_k]$: $CQ_k = |\{(i,j) \in E, 1 - \eta < \hat{a}_{ij}^{T_{k+1}} \cdot \hat{\alpha}_i^{T_k}/\hat{\alpha}_j^{T_k} < (1+\eta)^{-1}\}|/|E|$, where $\hat{a}_{ij}^{T_{k+1}}$ is the relative calibration estimated between fiber i and fiber j on the temporal window $]T_k, T_{k+1}]$, and η is an arbitrary threshold set to 0.1. The expression $\hat{a}_{ij}^{T_{k+1}} \cdot \hat{\alpha}_i^{T_k}/\hat{\alpha}_j^{T_k}$ is basically an unbiased estimation of the residuals of (6). CQ_k therefore provides an unbiased measure of how good we would be at reconstructing a flat image from a flat signal with the same physical properties as the observed medium. We then define the global calibration quality CQ of the whole video sequence as the average of static calibration quality values.

According to the results presented in Fig. 2 on the left, *online* blind calibration yields statistically higher calibration quality than *offline* calibration (95% confidence interval of $[0.13, 0.19]$ using paired difference t-test). Visual inspection of worst cases revealed that the sequences for which *offline* calibration performs similarly or slightly better contains either mostly noise, static fluorescence signal, non-sufficient moving fluorescent signal or only thin fluorescent structures. For all 89 sequences, the first *online* calibration succeeded after less than 30 seconds of moving fluorescent signal, a delay which is clearly compatible with

clinical practice. Regarding the sequences acquired with a probe of $n = 10,000$ fibers for example, computation time was less than 4 ms per frame in the buffer for performing all regressions, and less than 300 ms for system inversion.

We then propose to evaluate the impact of the calibration on image reconstruction by measuring the noise (static and dynamic) in a calibrated sequence. For this purpose, we compute the average of the (Delaunay-based) total variation [6] on all T frames of the sequence: $\sum_{t=1}^{T} \sum_{(i,j)\in E} |\hat{u}_i(t) - \hat{u}_j(t)|/(|E|T)$. Results show that the measured noise is statistically lower after *online* blind calibration than after *offline* calibration (95% confidence interval of $[-55, -33]$ using paired difference t-test). The improvement of image quality, from *offline*
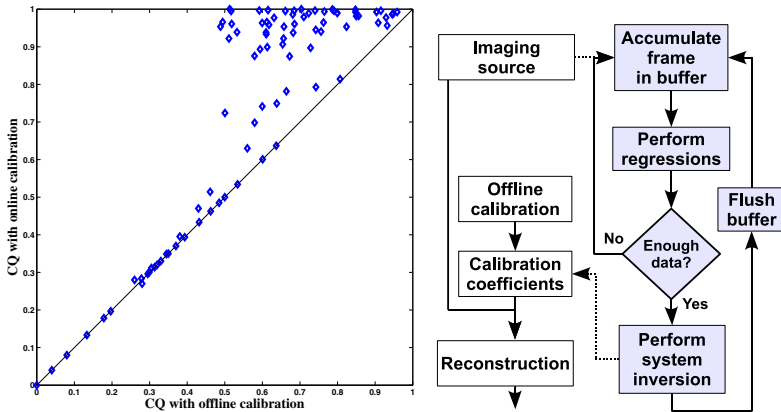
Fig. 2. Left: Calibration Quality (CQ) of 89 sequences processed by *online* versus *offline* calibration. Right: Pipeline of *online* blind calibration, highlighted in gray boxes. Dotted arrows show source input (top) and calibration update (bottom).
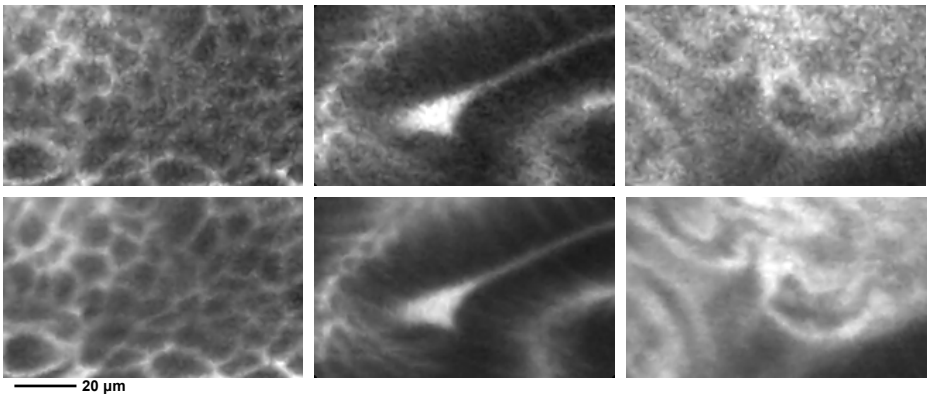
Fig. 3. Endomicroscopy image portions of Barrett's Esophagus, from 3 sequences processed by *offline* calibration (top) or *online* calibration (bottom). Bottom images appear much less noisy although no spatial smoothing is performed by *online* calibration.

to *online* calibration, can be qualitatively appreciated on still images presented in Fig. 3, and on two video sequences acquired on the bile duct and on the colon, available as Supplemental Material (http://youtu.be/1WYEQysDBqQ).

4   Conclusion

We have presented an original method for *online blind* calibration of non-uniform photodetectors, where only past signal is used to calibrate current signal in a transparent way to the user. By performing robust linear regressions and regularizing an ill-posed inverse problem, our method is able to handle photodetectors having both irregular spatial arrangement and individual slowly time-varying photometric transfer functions. Using a relatively large database of complete sequences acquired during clinical endomicroscopy procedures, we have demonstrated that *online* blind calibration statistically outperforms standard *offline* calibration. For future work, we plan to evaluate whether *online* blind calibration leads to higher perceived image quality and better diagnostic performance for the physicians.

References

1. Le Goualher, G., Perchant, A., Genet, M., Cavé, C., Viellerobe, B., Berier, F., Abrat, B., Ayache, N.: Towards Optical Biopsies with an Integrated Fibered Confocal Fluorescence Microscope. In: Barillot, C., Haynor, D.R., Hellier, P. (eds.) MICCAI 2004. LNCS, vol. 3217, pp. 761–768. Springer, Heidelberg (2004)
2. Winter, C., Rupp, S., Elter, M., Münzenmayer, C., Gerhäuser, H., Wittenberg, T.: Automatic adaptive enhancement for images obtained with fiberscopic endoscopes. IEEE Trans. Biomed. Eng. 53(10), 2035–2046 (2006)
3. Elter, M., Rupp, S., Winter, C.: Physically motivated reconstruction of fiberscopic images. In: Proc. ICPR 2006, vol. 3, pp. 599–602 (August 2006)
4. Kuhn, J.R., Lin, H., Loranz, D.: Gain calibrating nonuniform image-array data using only the image data. In: Proc. ASPC 1991, vol. 103, pp. 1097–1108 (1991)
5. Weiss, Y.: Deriving intrinsic images from image sequences. In: Proc. ICCV 2001, vol. 2, pp. 68–75 (2001)
6. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Phys. D 60(1-4), 259–268 (1992)
7. Pérez, P., Gangnet, M., Blake, A.: Poisson image editing. ACM Trans. Graph. 22(3), 313–318 (2003)
8. Agrawal, A., Chellappa, R., Raskar, R.: An algebraic approach to surface reconstruction from gradient fields. In: Proc. ICCV 2005, vol. 1, pp. 174–181 (2005)
9. Markovsky, I., Van Huffel, S.: Overview of total least-squares methods. Signal Processing 87(10), 2283–2302 (2007)
10. Berglund, A.J.: Nonexponential statistics of fluorescence photobleaching. J. Chem. Phys. 121(7), 2899–2903 (2004)

Author Index