Cristian A. Linte · Ziv Yaniv
Pascal Fallavollita (Eds.)

LNCS 9365

Augmented Environments for Computer-Assisted Interventions

10th International Workshop, AE-CAI 2015
Held in Conjunction with MICCAI 2015
Munich, Germany, October 9, 2015, Proceedings

Springer

EXTRAS ONLINE

Lecture Notes in Computer Science 9365

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

Editorial Board

More information about this series at http://www.springer.com/series/7412

Cristian A. Linte · Ziv Yaniv
Pascal Fallavollita (Eds.)

Augmented Environments for Computer-Assisted Interventions

10th International Workshop, AE-CAI 2015
Held in Conjunction with MICCAI 2015
Munich, Germany, October 9, 2015
Proceedings

Springer

Editors
Cristian A. Linte
Rochester Institute of Technology
Rochester, MN
USA

Pascal Fallavollita
Technical University of Munich
Munich
Germany

Ziv Yaniv
National Institutes of Health
Bethesda, MD
USA

and

TAJ Technologies Inc.
Mendota Heights, MN
USA

Videos to this book can be accessed at
http://www.springerimages.com/videos/978-3-319-24600-0

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Computer Science
ISBN 978-3-319-24600-0 ISBN 978-3-319-24601-7 (eBook)
DOI 10.1007/978-3-319-24601-7

Library of Congress Control Number: 2015950003

LNCS Sublibrary: SL6 – Image Processing, Computer Vision, Pattern Recognition, and Graphics

Springer Cham Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

Preface

Welcome to the proceedings of the 10th edition of the Augmented Reality for Computer-Assisted Interventions (AE-CAI) workshop. We are pleased to present the proceedings of this exciting workshop held in conjunction with MICCAI 2015 on October 9, 2015, in Munich, Germany.

The event was jointly organized by scientists from Rochester Institute of Technology (Rochester, NY, USA), National Institutes of Health and National Library of Medicine, and the Technical University of Munich (Munich, Germany). In addition, a Program Committee consisting of more than 70 international experts served as reviewers for the submitted papers.

The objective of the AE-CAI workshop was to attract scientific contributions that offer solutions to the technical problems in the area of augmented and virtual environments for computer-assisted interventions, and to provide a venue for dissemination of papers describing both complete systems and clinical applications. The community also encourages a broad interpretation of the field – from macroscopic to molecular imaging, passing the information on to scientists and engineers for the development of breakthrough therapeutics, diagnostics, and medical devices, which can then be seamlessly delivered back to patients. The workshop attracted researchers in computer science, biomedical engineering, computer vision, robotics, and medical imaging. This meeting featured a single track of oral and poster presentations showcasing original research engaged in the development of virtual and augmented environments for medical image visualization and image-guided interventions.

In addition to the proffered papers and posters, we were pleased to welcome as keynote speaker Dr. Simon Weidert (Ludwig Maximilian University, Munich, Germany) speaking on the development, clinical integration, and commercialization of computer-assisted surgery and simulation technology at the intersection of medicine and computer science.

AE-CAI 2015 attracted high-quality paper submissions from seven countries. The submissions were distributed for review to the Program Committee and each paper was evaluated, in a double-blind manner, by at least three experts, who provided detailed critiques and constructive comments to the authors and workshop editorial board. Based on the reviews, 15 papers were selected for oral and poster presentation and publication in these proceedings. The authors revised their submissions according to the reviewers' suggestions, and resubmitted their manuscripts, along with their response to reviewers, for a final review by the volume editors (to ensure that all reviewers' comments were properly addressed) prior to publication in this collection.

On behalf of the AE-CAI 2015 Organizing Committee, we would like to extend our sincere thanks to all Program Committee members for providing detailed and timely reviews of the submitted manuscripts. We also thank all authors, presenters, and attendees at AE-CAI 2015 for their scientific contribution, enthusiasm, and support.

We hope that you all will enjoy reading this volume and we look forward to your continuing support and participation in our future AE-CAI events.

August 2015

Cristian A. Linte
Ziv Yaniv
Pascal Fallavollita

Organization

Organizing Committee

Cristian A. Linte	Rochester Institute of Technology, Rochester, NY, USA
Ziv Yaniv	National Institutes of Health, Bethesda, MD, USA and TAJ Technologies Inc., Mendota Heights, MN, USA
Pascal Fallavollita	Technical University of Munich, Munich, Germany

Program Committee

Takehiro Ando	The University of Tokyo, Japan
Leon Axel	New York University Medical Center, USA
Adrien Bartoli	Université d'Auvergne, France
John Baxter	Robarts Research Institute, Canada
Marie-Odile Berger	Inria, France
Wolfgang Birkfellner	Medical University of Vienna, Austria
Yiyu Cai	Nanyang Technological University, Singapore
Xinjian Chen	Soochow University, China
Chung-Ming Chen	National Taiwan University, Taiwan
Elvis Chen	Robarts Research Institute, Canada
Kevin Cleary	Children's National Medical Center, USA
Louis Collins	McGill University, Canada
Fabrizio Cutolo	EndoCAS Center, University of Pisa, Italy
Simon Drouin	McGill University, Canada
Eddie Edwards	Imperial College London, UK
Gary Egan	University of Melbourne, Australia
Pascal Fallavollita	Technical University of Munich, Germany
Yong Fan	Chinese Academy of Sciences, China
Gabor Fichtinger	Queen's University, Canada
Michael Figl	Medical University of Vienna, Austria
Kenko Fujii	Imperial College London, UK
John Galeotti	Carnegie Mellon University, USA
James Gee	University of Pennsylvania, USA
Stamatia Giannarou	Imperial College London, UK
Ali Gooya	University of Pennsylvania, USA
Lixu Gu	Shanghai Jiaotong University, China
Makoto Hashizume	Kyushu University, Japan
David Hawkes	University College London, UK
David Holmes III	Mayo Clinic, USA

Jaesung Hong	Daegu Gyeongbuk Institute of Science and Technology, Korea
Robert Howe	Harvard University, USA
Pierre Jannin	Université de Rennes I, France
Bernhard Kainz	Imperial College London, UK
Ali Kamen	Siemens Corporate Research, USA
Ron Kikinis	Harvard University, USA
Jan Klein	Fraunhofer MEVIS, Germany
David Kwartowitz	Clemson University, USA
Rudy Lapeer	University of East Anglia, UK
Su-Lin Lee	Imperial College London, UK
Ming Li	National Institutes of Health, USA
Tianming Liu	University of Georgia, USA
Jimmy Liu	Agency for Science, Technology and Research, Singapore
Gian-Luca Mariottini	University Texas at Arlington, USA
John Moore	Roberts Research Institute, Canada
Kensaku Mori	Nagoya University, Japan
Ryoichi Nakamura	Chiba University, Japan
Nassir Navab	Technical University of Munich, Germany
Philip Pratt	Imperial College London, UK
Maryam Rettmann	Mayo Clinic, USA
Jannick Rolland	University of Rochester, USA
Ichiro Sakuma	The University of Tokyo, Japan
Yoshinobu Sato	Nara Institute of Science and Technology, Japan
Pengcheng Shi	Rochester Institute of Technology, USA
Amber Simpson	Memorial Sloan-Kettering, USA
George Stetten	University of Pittsburgh and Carnegie Mellon University, USA
Danail Stoyanov	University College London, UK
Russell Taylor	Johns Hopkins University, USA
Tamas Ungi	Queen's University, Canada
Theo Van Walsum	Erasmus Medical Center, The Netherlands
Kirby Vosburgh	Harvard University, USA
Guangzhi Wang	Tsinghua University, China
Jaw-Lin Wang	National Taiwan University, Taiwan
Junchen Wang	The University of Tokyo, Japan
Stefan Wesarg	Fraunhofer University, Germany
Kelvin Wong	Methodist Hospital - Weill Cornell Medical College, USA
Jue Wu	University of Pennsylvania, USA
Zhong Xue	Weill Cornell Medical College, USA
Yasushi Yamauchi	Toyo University, Japan
Guang-Zhong Yang	Imperial College London, UK
Jong Chul Ye	Korea Advanced Institute of Science and Technology, Korea
Bo Zheng	The University of Tokyo, Japan

Contents

Ultrasound-Guided Navigation System for Orthognathic Surgery

Beatriz Paniagua[1(✉)], Dženan Zukic[2], Ricardo Ortiz[2], Stephen Aylward[2], Brent Golden[3], Tung Nguyen[3], and Andinet Enquobahrie[2]

[1] Departments of Psychiatry, Computer Science and Orthodontics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
beatriz_paniagua@unc.edu
[2] Kitware Inc., Carborro, NC, USA
{dzenan.zukic, ricardo.ortiz, stephen.aylward, andinet.enqu}@kitware.com
[3] School of Dentistry, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
{brent_golden, tung_nguyen}@unc.edu

Abstract. Around 1–2 % of the US population has craniofacial deformities severe enough to be disabling and stigmatizing, and could benefit from orthognathic surgery. This surgery involves repositioning the jaws, due to the unique features of each patient's teeth, jaws, and joint. Approximately 20 % of patients who had mandibular advancement surgery experience moderate relapse 1–5 years after surgery. We believe ultrasound is a promising imaging technology for orthognathic surgery guidance that can assist surgeons to visualize the condyle/ramus segment in order to guide it into its pre-surgical, biologically stable position. This paper explores the role of 3D ultrasound imaging as a real-time surgical guidance to improve treatment outcomes for orthognathic surgery. This paper shows our work designing a 3D ultrasound volume reconstruction system and our results demonstrating its ability to capture the bony structures of the mandible, compared with those structures reconstructed from pre-surgical Cone Beam Computed Tomography (CBCT).

Keywords: Ultrasound · Computer guided interventions · Orthognathic surgery · Cone beam computed tomography

1 Introduction

Around 1–2 % of the US population has craniofacial deformities severe enough to be disabling and stigmatizing, and could benefit from reconstructive surgery [1]. If left untreated, craniofacial deformities can produce physical problems such as speech, respiratory and masticatory problems, and psychosocial problems such as teasing, stereotyping or bullying [2, 3]. Correction of these deformities using orthognathic surgery involves precise repositioning of the jaws, due to the unique features of each patient's teeth, jaws, and joint.

C.A. Linte et al. (Eds.): AE-CAI 2015, LNCS 9365, pp. 1–10, 2015.
DOI: 10.1007/978-3-319-24601-7_1

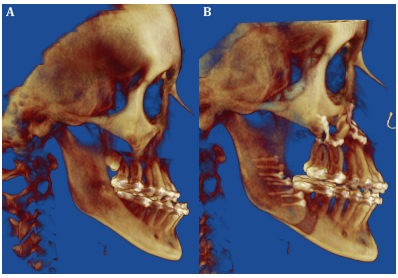


Fig. 1. A patient with a Class III dentofacial deformities requiring orthognathic surgery. Pre-operative CBCT (A) and post-operative CBCT (B) are shown.

Surgical techniques have improved over the past 50 years decreasing post-operative complications. The introduction of 3D imaging techniques [4–7] and 3D splint fabrication [8] have improved aesthetic and functional outcomes. However, skeletal surgical relapse, which relates to the return of bony tissues to their pre-surgical state, still remains a problem. Approximately 20 % of patients who had mandibular advancement surgery experience moderate relapse 1–5 years after surgery [9]. A major factor in post-operative relapse lies with the inability to precisely relocate the temporomandibular joint in the exact pre-operative position.

Ultrasound (US) imaging has attracted the attention of researchers in the dental and cranio-maxillo facial surgery fields since it is a non-invasive, non-ionizing imaging technology that is inexpensive and readily available. Recent studies show that it has been used in the assessment of facial soft tissues before and after orthognathic surgery [10]. Ultrasound has been also employed to capture position of the temporomandibular joint (TMJ) as well as its dynamic properties during movement [11–14].

We believe ultrasound is a promising imaging technology for orthognathic surgery guidance. This imaging technology, combined with real-time 3D registration techniques can assist Oral Maxillofacial Surgeons during surgery to visualize the condyle/ramus segment in order to guide it into its pre-surgical, biologically stable position. Having the condyle/ramus segment is in its biologically stable position can potentially reduce surgical relapse and improve treatment outcomes.

This paper presents the preliminary results we obtained by exploring the role of real-time surgical guidance using 3D ultrasound imaging to improve treatment outcomes for orthognathic surgery.

2 Materials

2.1 CBCT

Cone-beam CT (CBCT) images are acquired for comparison with ultrasound reconstructions using the New-Tom 3G scanner (AFP Imaging, Elmsford, NY). The imaging protocol involves a 36-second head CBCT scan with a 12-inch field of view. After the

images are acquired, the next step is image segmentation, where the anatomical structures of interest are identified and delineated in the CBCT image. In orthognathic surgery, the goal of segmentation is to obtain a 3D representation (surface model) of the hard tissue that is essential for virtual planning. Currently available 3D image analysis software tools utilize manual, semi-automatic and/or fully automatic segmentation techniques (Dolphin [15], 3DMD Vultus [16] and Maxillim [17]). Automatic segmentation, which relies on greyscale thresholding algorithms, does not offer the best results for virtual surgery planning, since different bony structures in the mandible have different bone density and thickness. The condyles often have thin, less dense bone compared the ramus or body of the mandible. Therefore an automated threshold level that accurately reproduces the 3D surface structure of the ramus, will not capture the anatomic details of the condyles or vice versa.

To best capture all the relevant structures for orthognathic surgery, we used the semi-automatic segmentation procedures provided by OrthognathicTrac. **OrthognathicTrac** is a 3D Slicer based GUI application that is customized for surgical planning of orthognathic surgery (Fig. 1). OrthognathicTrac is designed for ultrasound 3D reconstruction and CBCT segmentation, and as a platform to easily deploy new algorithms. After obtaining the segmentation result, manual post-processing is necessary to remove common artifacts resulting from metallic elements. After segmentation of the anatomical structures of interest, OrthognathicTrac allows a 3D display of the anatomical areas directly from the volume data.

2.2 Ultrasound

For 3D ultrasound volume acquisition, we built a tracked 3D freehand ultrasound image acquisition system. This acquisition system consists of multiple components. The major components are the 2D ultrasound probe, a tracking device and tracked markers, and a computer for image processing, storage and display (Fig. 2). In this project, we used the MicronTracker H40 (passive optical tracker from ClaroNav, 1140 Sheppard Avenue West, Unit 10, Toronto, M3 K 2A2 Ontario, Canada) and TeleMed

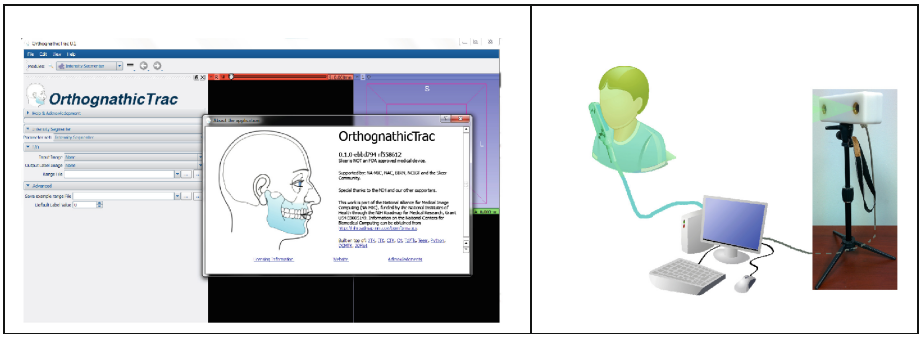


Fig. 2. Left: OrthognaticTrac graphic user interface. Right: A 3D tracked ultrasound acquisition set-up consist of computer, optical tracker and ultrasound probe.

LogicScan 128 ultrasound scanner with probe model LV7.5/60/128Z-2 (60 mm long 8 MHz linear probe from Telemed, Dariaus ir Gireno str. 42, Vilnius LT-02189, Lithuania).

The goal is to acquire multiple 2D ultrasound scans and their corresponding tracking information to create a 3D ultrasound volume. The 3D Ultrasound volume acquisition involves two phases (1) Calibration and (2) Data acquisition and reconstruction.

Calibration: US system calibration involves 4 steps i.e. Stylus calibration, phantom registration, temporal calibration and then spatial calibration, and it needs to be performed once. During the ***stylus calibration*** step, we compute the coordinate transformation from the stylus tip to the tracker fiducials. For this we use a pivot-calibration algorithm, that first places the stylus in a position that allows the tip of the needle to remain stationary and for all of tracker fiducials to be seen by the camera, then swivels the stylus (tip remains stationary) until enough points are collected and finally runs a pivot calibration algorithm to compute the transformation between stylus and tracker fiducials. The ***phantom registration*** step's goal is to compute the coordinate transformation between the tracker and the phantom model, by identifying pre-designated reference points on the phantom and record the location using the stylus and then running landmark registration to compute the transformation. The ***temporal calibration*** step computes the time lag of the US probe's tracker stream relative to the US video stream. Finally, the ***spatial calibration*** step determines the coordinate transformation between the ultrasound probe and the image, by scanning the calibration phantom in a water bath. For this, we use N-wire calibration phantom in which threads forming a shape like uppercase letter N are attached to a 3D printed plastic template.

The probe calibration and 3-D volume construction software were developed using the Public software Library for Ultrasound Imaging research (PLUS), an open source toolkit for translational research of ultrasound-guided intervention systems [18].

Data Acquisition: Once the system is calibrated, it is possible to start acquiring 2D ultrasound scans and associated tracking information in order to run the 3D Volume reconstruction algorithm. To acquire 3D ultrasound scans, we first sweep the

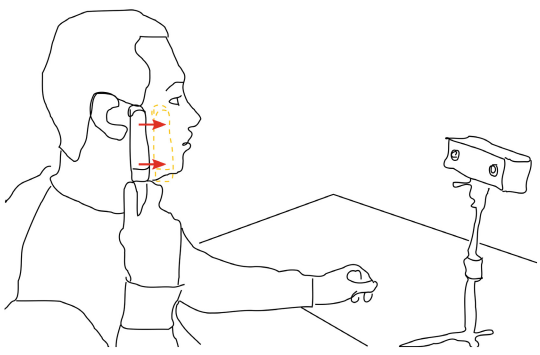


Fig. 3. Acquisition pose for 3D ultrasound data

ultrasound probe in the posterior-anterior direction in the area of mandibular ramus (see Fig. 3) to acquire 2D images and its associated tracking information. Starting with an empty volume, slices are inserted into their correct spatial position and orientation. At the end, a hole filling step [18] is performed to fill minor gaps due to jerky movement of the probe or similar issues. This ultrasound volume is what we will use for intra-surgical registration.

3 Methods

For the pilot work presented in this paper, two volunteers (RO and DZ) were recruited, and CBCT and 3D ultrasound scans were acquired. Approval was obtained from the University of North Carolina Institutional Review Board (study #13-4096). From those scans, we generated segmentations and 3D surface models to register 3D ultrasound volumes for left and right mandibular ramus and CBCT data using functions available in OrthognathicTrac.

3.1 Surface Model Generation

Surface models representing the bony anatomic structures of the mandible are generated from 3D ultrasound and CBCT volumes. **CBCT** is the image modality that is commonly acquired as part of the pre-surgical clinical protocol for malocclusion 3D diagnosis and treatment planning. We consider CBCT as the off-line modality that can be prepared carefully for ultrasound surgical guidance during orthognathic surgery. The first step to generate surface models from CBCT images is to perform image segmentation, where the anatomical structures of interest are identified and delineated in the image. For orthognathic surgery the goal of CBCT segmentation is to obtain a 3D surface model of the bony structures that is usable for computer-assisted interventions.

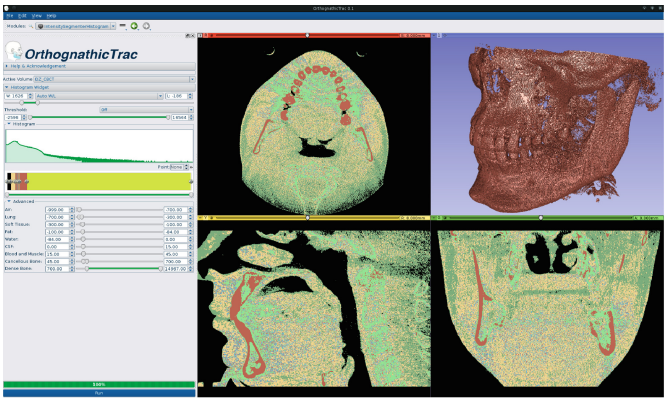


Fig. 4. OrthognathicTrac semi-automatic segmentation.

Key structures include the posterior and inferior borders of the mandible, the zygomatic bone, the external auditory meatus, and the mastoid process. OrthognathicTrac offers a semi-automatic, multi-threshold segmentation methods (see Fig. 4) that produce more accurate 3D representations of the anatomic structures of interests. Many commercial automatic segmentation software rely on a single threshold algorithm which cannot accurately reproduce anatomical structures of varying bone density objects such as the condyle, ramus, and body of the mandible.

Ultrasound surface generation is done by first generating a segmentation using a thresholding-based method. The thresholding-based method is computationally very fast and it requires minimal user interaction. Then, a de-noising algorithm consisting in a closing operation with a structuring element of $1 \times 1 \times 1$ voxel and a median filter of kernel $2 \times 2 \times 2$ voxels is performed. Other high intensity structures captured in the 3D ultrasound such as the masseter muscle are manually removed from the label map. Surface models for both image modalities are then generated using a marching cubes algorithm [19].

3.2 Surface Models Registration

Landmark registration was used for initial alignment of the 3D ultrasound and CBCT scans. Corresponding anatomical (landmark) points are identified in the parts of the mandible that are visible in both mandibular models such as the retromolar fossa. Landmark points were usually between 4–5 points, but not less than three, to be able to capture the best transformation between both models. Since high-frequency linear probe does not introduce any geometric distortion to captured structures, a rigid body transform can accurately register 3D surface models from 3D ultrasound and CBCT. Landmark-based registration was used instead of surface registration since US reconstruction does not reconstruct bony structures as a whole, like it happens in CBCT, therefore a point-cloud surface registration would have trouble finding the best match between both surface models. Landmark-based registration is computationally very fast.

3.3 Registration of Pre-surgical Off-line CBCT to On-line Ultrasound for Computer-Assisted Surgical Guidance

The goal of this step is to implement intra-operative 3D ultrasound to CBCT registration algorithms to combine pre-surgical CBCT generated surfaces (zygomatic bone, auditory meatus and mastoid process) on to the ultrasound image for real-time feedback. OrthognathicTrac allows applying any previously calculated rigid body transform to any source data. In order to register 3D ultrasound to CBCT volumes we apply the previously calculated transform (see Sect. 3.2) to the 3D ultrasound volumes. Error distances between the two registered surfaces were calculated using VAM (Vectra Scientific, Atlanta, GA)

4 Results

This study demonstrates 3D ultrasound can accurately capture the outer surfaces of the mandibular bones. Figure 5 shows the registered superimposition of the 3D US surface (red) to the CBCT mandibular surfaces (tan). The shape and contour of the US generated bony surfaces aligned well with that of the CBCT generated surfaces.

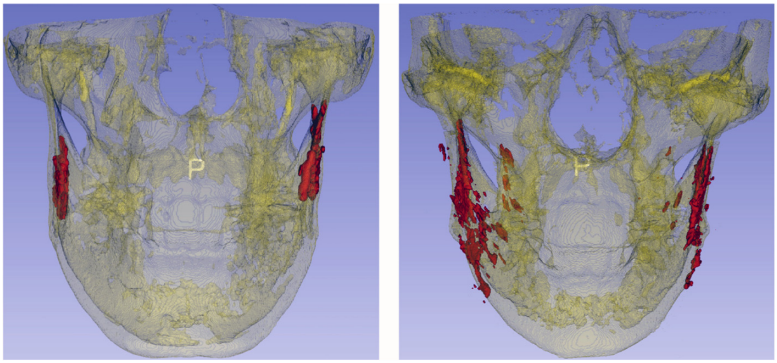


Fig. 5. 3D Surface model registration: 3D ultrasound reconstruction (red) and CBCT (yellow) surface models show that ultrasound is able to accurately capture the outer layer of bone in the mandibular ramus (Color figure online).

In addition, error distances between the US and CBCT surfaces were calculated. On average, the Root Mean Square (RMS) error between the US surfaces and CBCT surfaces was 0.51 mm, with a mean total surface error of 0.45 mm (Table 1).

Table 1. Descriptive statistics of error distances between registered US surfaces and CBCT surfaces.

Patient	Max	RMS	Mean	StdDev
RO Left	1.475	0.71	0.654	0.29
RO Right	1.293	0.37	0.311	0.214
DZ Left	1.205	0.49	0.401	0.29
DZ Right	1.237	0.50	0.434	0.60
Absolute Error	1.3025	0.5175	0.45	0.3485

Our data indicates that 3D ultrasound and CBCT volume registration show accurate registration results, and that precise outer layers of bone in the mandibular ramus can be visualized in 3D ultrasound (Fig. 6).

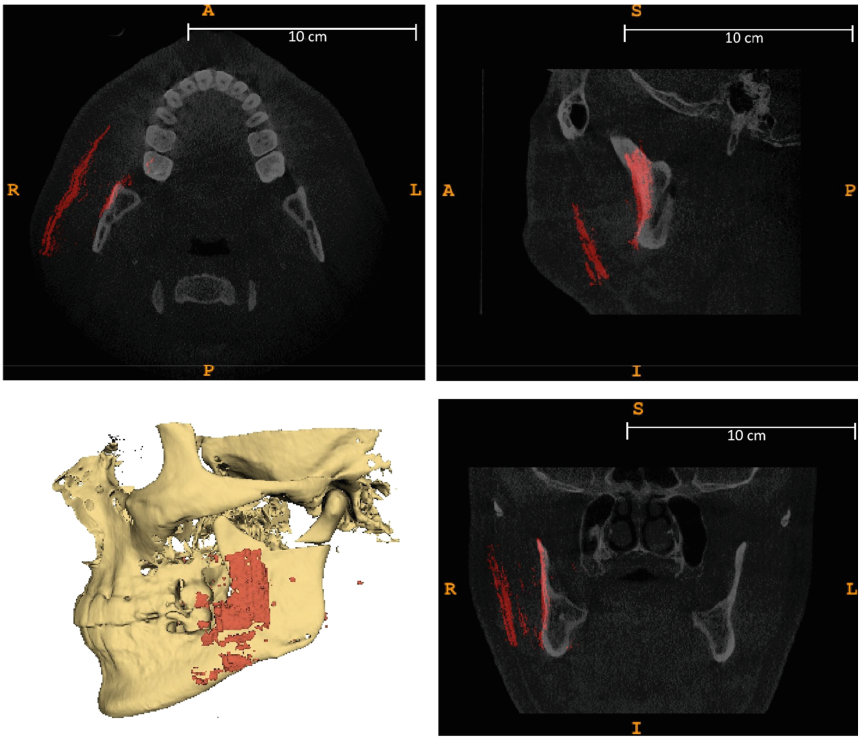


Fig. 6. Coronal view of CBCT (grayscale) and 3D ultrasound (red semitransparent) volume overlay. Areas of high ultrasound reflection depicted in red. Both the masseter muscle and the mandibular ramus surface area can be clearly visualized in both modalities. In slice cross-sections, deeper layers of skin are visible in the ultrasound scan (red semitransparent). The surface layers of skin were removed during US acquisition by 2D image cropping (Color figure online).

5 Conclusions

This work demonstrates that it is possible to reconstruct high-density bone in the mandibular ramus from US imaging, and register it to CBCT. This shows the potential of US as non-invasive, accurate, low-cost real-time imaging modality to help confirm the position of the condyle within the mandibular fossa by rendering the position of the connected bony structures (ramus, gonial angle, mandibular notch) and reduce surgical relapse. As part of this work, we have also developed **OrthognathicTrac**, which is a platform to develop, prototype, and test algorithms for orthognathic surgery planning, navigation and visualization. Setting up the US pipeline for 3D ultrasound reconstruction was not trivial and several problems related with US parameters and frequency were encountered and solved in the course of this research. Also, and even with the current data acquisition setup, the calibration requirements and line of sight issues are problems to resolve before the proposed framework can be integrated fully in the

operating room. To address these problems, we intend to explore tracker-less 3D ultrasound acquisition systems as part of the future work. Inter-rater reliability of US to CBCT model registration using landmarks will need to be also studied in the future.

Future work contemplates evaluating the effectiveness and feasibility of the algorithms implemented in the OrthognathicTrac prototype using a head phantom and mimicking operating room conditions, as well as improving the current processing time of 5 min and test its clinical viability via mock up surgical case.

Acknowledgements. Research reported in this publication was supported by the National Institute of Dental & Craniofacial Research of the National Institutes of Health under Award Number R43DE024334. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

1. Bailey, L.J., Duong, H.L., Proffit, W.R.: Surgical Class III treatment: long-term stability and patient perceptions of treatment outcome. Int. J. Adult Orthodon. Orthognath. Surg. **13**, 35–44 (1998)
2. Rankin, M., Borah, G.L.: Perceived functional impact of abnormal facial appearance. Plast. Reconstr. Surg. 111, 2140–6; discussion 2147–8 (2003)
3. Phillips, C., Bennett, M.E., Broder, H.L.: Dentofacial disharmony: psychological status of patients seeking treatment consultation. Angle Orthod. **68**, 547–556 (1998)
4. Xia, J., Ip, H.H., Samman, N., Wang, D., Kot, C.S., Yeung, R.W., Tideman, H.: Computer-assisted three-dimensional surgical planning and simulation: 3D virtual osteotomy. Int. J. Oral Maxillofac. Surg. **29**, 11–17 (2000)
5. Xia, J.J., Gateno, J., Teichgraeber, J.F., Christensen, A.M., Lasky, R.E., Lemoine, J.J., Liebschner, M.A.K.: Accuracy of the computer-aided surgical simulation (CASS) system in the treatment of patients with complex craniomaxillofacial deformity: A pilot study. J. Oral Maxillofac. Surg. **65**, 248–254 (2007)
6. Cevidanes, L.H.C., Tucker, S., Styner, M., Kim, H., Chapuis, J., Reyes, M., Proffit, W., Turvey, T., Jaskolka, M.: Three-dimensional surgical simulation. Am. J. Orthod. Dentofacial Orthop. **138**, 361–371 (2010)
7. Hsu, S.S.-P., Gateno, J., Bell, R.B., Hirsch, D.L., Markiewicz, M.R., Teichgraeber, J.F., Zhou, X., Xia, J.J.: Accuracy of a computer-aided surgical simulation protocol for orthognathic surgery: a prospective multicenter study. J. Oral Maxillofac. Surg. **71**, 128–142 (2013)
8. Bell, R.B.: Computer planning and intraoperative navigation in orthognathic surgery. J. Oral Maxillofac. Surg. **69**, 592–605 (2011)
9. Proffit, W.R., Fields, H.W., Moray, L.J.: Prevalence of malocclusion and orthodontic treatment need in the United States: estimates from the NHANES III survey. Int. J. Adult Orthodon. Orthognath. Surg. **13**, 97–106 (1998)
10. Trawitzki, L.V.V., Dantas, R.O., Elias-Júnior, J., Mello-Filho, F.V.: Masseter muscle thickness three years after surgical correction of class III dentofacial deformity. Arch. Oral Biol. **56**, 799–803 (2011)
11. Stefanoff, V., Hausamen, J.E., van den Berghe, P.: Ultrasound imaging of the TMJ disc in asymptomatic volunteers. Preliminary Rep. J. Craniomaxillofac. Surg. **20**, 337–340 (1992)

12. Landes, C.A., Sterz, M.: Evaluation of condylar translation by sonography versus axiography in orthognathic surgery patients. J. Oral Maxillofac. Surg. **61**, 1410–1417 (2003)
13. Gateno, J., Miloro, M., Hendler, B.H., Horrow, M.: The use of ultrasound to determine the position of the mandibular condyle. J. Oral Maxillofac. Surg. 51, 1081–6; discussion 1086–7 (1993)
14. Hayashi, T., Ito, J., Koyama, J., Yamada, K.: The accuracy of sonography for evaluation of internal derangement of the temporomandibular joint in asymptomatic elementary school children: comparison with MR and CT. AJNR Am. J. Neuroradiol. **22**, 728–734 (2001)
15. Dolphin Imaging
16. 3dMD: 3DMD. http://www.3dmd.com/
17. Maxilim Surgical Planning Software. http://www.dentalcompare.com/4885-Implant-Surgical-Guide-Systems/41444-Maxilim-Maxillofacial-Surgery-Planning-Software/
18. PLUS library. http://perk-software.cs.queensu.ca/plus/doc/nightly/user/index.html
19. Lorensen, W.E., Cline, H.E.: Marching cubes: A high resolution 3D surface construction algorithm. ACM SIGGRAPH Comput. Graph. **21**, 163–169 (1987)

Augmented Reality Ultrasound Guidance for Central Line Procedures: Preliminary Results

Golafsoun Ameri[1,2(✉)], John S.H. Baxter[1,2], A. Jonathan McLeod[1,2], Terry M. Peters[1,2], and Elvis C.S. Chen[1]

[1] Robarts Research Institute, London, ON, Canada
chene@robarts.ca
[2] Biomedical Engineering Graduate Program, Western University, London, ON, Canada
{gameri,jbaxter,jmcleod,tpeters}@robarts.ca

Abstract. Central line procedures are interventions in which a needle is placed in the jugular vein in the patient's neck inferior to the carotid bifurcation. In these procedures, avoiding the puncture of the carotid artery is of upmost importance as it can cause severe neurological consequences or death. Often, these procedures are performed under ultrasound guidance, meaning that a linear ultrasound probe is held to the patient's neck in which the interventionalist can visualize both the carotid artery and jugular vein. However, due to the geometry of the interventional scene, the needle must be placed out-of-plane with the ultrasound and the needle cannot be fully visualized, only a cross-section thereof. This lack of visualization can lead to issues gaging the correct penetration depth. This paper presents preliminary results on an augmented reality (AR) needle guidance system in which a tracked needle and ultrasound fan are simultaneously visualized in their entirety. This AR guidance system is compared against traditional ultrasound-only guidance on a neck phantom. The use of the AR system significantly reduces the intervention time (average decrease of 3.51 ± 1.44 s) and normalized path length (average decrease of $150 \pm 40\,\%$) implying that the use of such as system makes the procedure easier for the interventionalist ($n = 36$, $p \leq 0.05$). This AR system has gained regulatory approval and is scheduled for clinical trials in humans.

Keywords: Central line procedure · Needle guidance · Ultrasound-guided interventions · Augmented reality

1 Introduction

Central venous cannulation, predominantly in the internal jugular, femoral, and subclavian veins, is a widely performed procedure in current medical practice often used in intensive care units and operating rooms. Among the main central venous catheter sites, the right internal jugular vein (IJV) cannulation is perhaps the most popular method providing access to the deep venous system

C.A. Linte et al. (Eds.): AE-CAI 2015, LNCS 9365, pp. 11–20, 2015.
DOI: 10.1007/978-3-319-24601-7_2

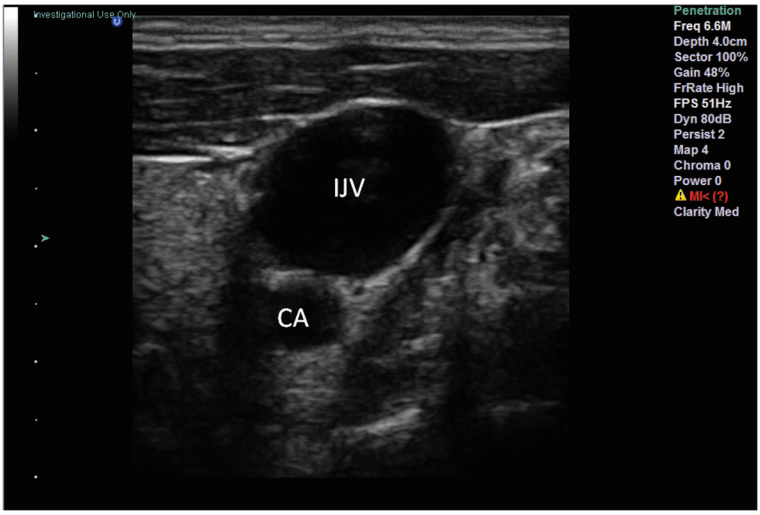


Fig. 1. Ultrasound images of a healthy volunteer. The image has been labeled with the carotid artery (CA) and internal jugular vein (IJV) for reference.

for a large number of applications including hemodynamic monitoring with pulmonary artery catheter, access for prolonged intravenous therapy, and access for endomyocardial biopsy [4]. Over 5 million central venous catheters are inserted annually in the US alone with an overall complication rate of 15 % [7]. Inadvertent carotid artery (CA) puncture can occure either by needle placement directly into the artery lumen, or after the cannulating needle transverses the IJV [8]. Severe neurological problems and life thretening complications may arise during central venous needling and catheterization due to arterial puncture and hematoma [5,7,9]. A safe needle and catheter placement is of utmost importance in central venous catheterization procedures.

Commonly, the IJV is identified using an external anatomical landmark-guided technique and then the introducer needle and subsequently the catheter are inserted blindly. However, an aberrant anatomical position of the IJV (in 8.5 % of patients) would make this technique unreliable even in the hands of the experienced interventionalists [4]. The reported failure rate to access to the IJV ranges from 7 %–19.4 % depending, in part, on the operator's experience [4].

Ultrasound (US) has been used off-line to localize the CA, the IJV, and the insertion site prior to cannulation. An ultrasound image of a healthy volunteer is shown in Fig. 1. The pulsation of the CA is visible under real-time US and has been suggested to be used to distinguish it from the IJV [8]. Real-time US has also been employed to guide needle placement inside the IJV. Ultrasound guidance during internal jugular catheterization has been shown to reduce the insertion time, the rates of unsuccessful catheterization, CA puncture, and hematoma formation [7].

The IJV is at a depth of 1.0–1.5 cm beneath the skin surface in most patients [2] and can be easily visualized in US using a 5–10 MHz linear transducer [7]. However, complications in central venous cannulation still arise in spite of real-time US guidance [3]. Blaivas *et al.* [3] reported that 68 % of emergency medicine residents, undergone 2-day US-guided vascular access training, inadvertently penetrated the posterior wall of the IJV during US-guided cannulation in a phantom study. The main reason for the failure of the US guidance in preventing any complication in IJV cannulation is that the US guidance is provided using a transverse approach, where the US image plane is perpendicular to the length of the vein [3]. As a result, only the cross section of the venous lumen is seen in the US image. In addition, the needle is inserted out-of-plane with the ultrasound and as a result only its cross section appears in the US image, as a bright spot, making it difficult to distingush the needle tip from needle shaft. Therefore, even under US guidance, issues arise regarding the correct needle placement and evaluation of the penetration depth.

The purpose of this work is to present a novel guidance system, which has been approved for clinical evaluation for the IJV cannulation procedure to facilitate needle navigation and placement in the IJV and improve the overall safety of this procedure. The proposed guidance system consists of an augmented reality (AR) environment, in which a tracked needle, its trajectory, and real-time US images are simultaneously visualized in the same coordinate system. The AR guidance system is compared against traditional US-only guidance on a neck phantom in a user study. This system has the potential to improve accuracy and safety of the central venous cannulation procedure in a clinical setting and has received regulatory approval for clinical trials in humans.

2 Methods

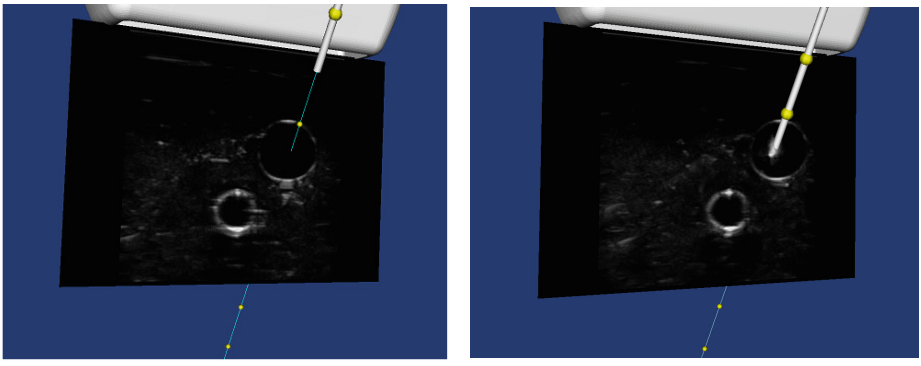
2.1 Augmented Reality Guidance System

An AR guidance platform was developed, which allowed for the US image and a virtual representation of the needle to be displayed simultaneously in the same coordinate frame. Ultrasound images were acquired using a SonixTouch US scanner and a magnetically tracked L14-5/38 ultrasound probe (Ultrasonix, Analogic Corporation, MA, USA). An Aurora magnetic tracking system (NDI, Waterloo, ON, Canada) was used to determine the physical location and orientation of the ultrasound probe and needle for use in the visualization. The US probe was calibrated using a line fiducial as described in [1]. The needle used was an 18G Aurora needle, which has a 5DOF magnetic sensor integrated into its stylet's tip. To improve the usability of the AR system, the virtual needle was equipped with a representation of its trajectory as well as a sequence of uniformly spaced 1 cm markings along the needle and its trajectory, giving the user a rapid understanding of the depth of penetration beyond, or remaining depth to, the ultrasound plane. The open source Atamai Image Guided Surgery (AIGS) library[1] was used

[1] https://github.com/awiles/AIGS.

for communication with tracking systems. The Visualization Toolkit (VTK)[2] and QT were used for visualization and user interface[3], respectively. The AR system visualization is shown in Fig. 2.

The developed needle guidance system is similar to the commercial SonixGPS (Ultrasonix, Analogic Corporation, MA, USA) system. However, it uses an Aurora tracking system instead of Ascension, which is used in SonixGPS. In addition, our custom AR system for needle visualization is substantially different from that used in SonixGPS and provides a 3D visualization environment. The advantage of our visualization system is that the needle can be guided and visualized in any orientation relative to the US probe (in-plane and out-of-plane), and the user can adjust the vantage point of the virtual camera to any arbitrary angle as needed. As SonixGPS is not widely available across all institutions/hospital, and the blind/US-guided technique is the current gold-standard, we conduct our study to illustrate the efficacy of the developed AR system compared to the clinical gold-standard.



(a) Needle prior to puncturing the IJV (b) Needle after positioning in the IJV

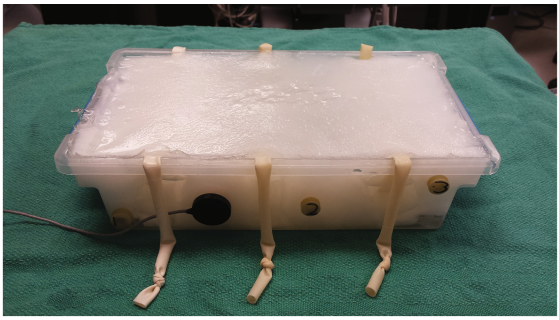
Fig. 2. Augmented reality needle guidance platform.

2.2 Phantom Construction

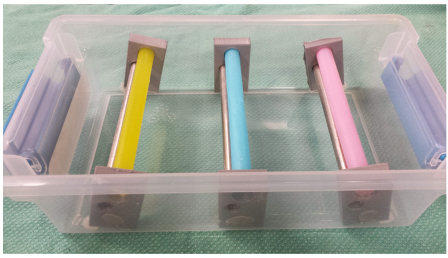
A phantom was designed to represent three IJV and CA sets for a user study, as shown in Fig. 3a. The IJV and the CA were simulated as hollow tubes inside a tissue mimicking polyvinyl alcohol (PVA-C) block (10 % PVA-C). In order to create the hollow tubes, plastic straws and metal tubes of different sizes were used.

The mean diameters in adults of the right IJV and the common CA is 14 ± 5 mm and 6.5 ± 1.0 mm, respectively [6,10]. The IJV is usually positioned in laterally and anteriorly to the CA at a depth of 1 to 1.5 cm below the skin [4].

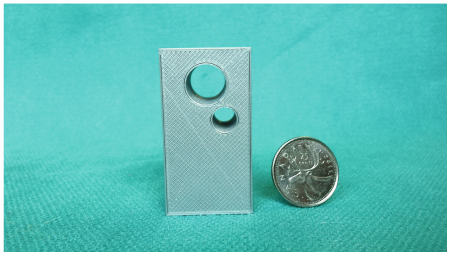
[2] http://www.vtk.org/.
[3] http://www.qt.io/.



(a) User Study Phantom consisting of a PVA-C block with water-filled tubes simulating the CA and IJV



(b) Phantom prior to introduction of PVA-C, showing three CA/IJV vascular combinations



(c) Rapid prototyped support block to keep the straws and tubes in place

Fig. 3. Phantom employed in user study.

In our phantom, the diameter of the hollow tubes representing the IJV and the CA were 12.3 mm and 8 mm, respectively. Three supporting blocks, Fig. 3c, were placed at three different angles along the length of the container on each side in order to account for anatomical variations and reduce training effects. Straws and metal tubes were placed inside the holes of the support block as shown in Fig. 3b. The plastic container was then filled with PVA-C and subjected to two freezing-thaw cycles, after which the straws were removed. A penrose tube filled with water was passed through the smaller tube so that it could be pulsed, by squeezing one end periodically, in order to create a more realistic representation of the CA as visualized in US. An Aurora 6DOF magnetic sensor, 25 mm Disc, (NDI, Waterloo, ON, Canada) was placed on the phantom box as a tracking reference to accommodate for any overall shift. An ultrasound image of the phantom is shown in Fig. 4.

2.3 User Study

Eighteen novice participants were recruited. After a brief training phase (approximately 20 min) in which the participants could familiarize themselves with both the US and AR guidance systems, the participants performed a set of two central

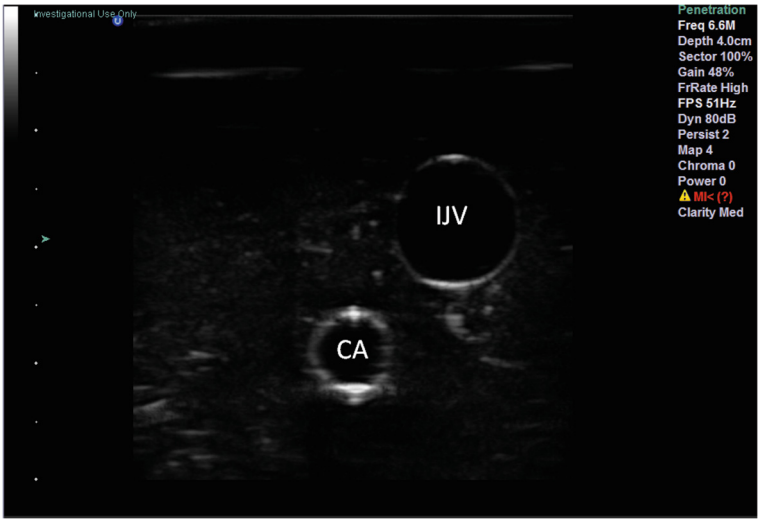


Fig. 4. Ultrasound images of the phantom. In comparison to Fig. 1, the phantom provides a realistic US image of the IJV and CA.

line procedures, one with AR guidance and one with US only. The order in which these were performed was randomized to control for the training effect. The task was defined as complete when the participant reported the needle to be in the IJV in the phantom. After a period of one week, each participant performed an additional set of central line procedures and the opposite order was used. In this study, the desired target for needle placement is any location within the phantom IJV. This is similar to the clinical central venous cannulation procedure in which any location along the IJV inferior to the carotid bifurcation is considered adequate for central line insertion.

For each procedure, two performance metrics were collected: procedure time and needle path tortuosity. The time taken to deliver the procedure was measured directly by our tracking software. Lower procedure times are desirable to minimize patient discomfort and to improve clinical workflow. The tortuosity of the needle path, how much the needle path deviates from a straight line during the procedure, is calculated using the normalized path length. The normalized path length is the ratio of the actual travelled path, divided by the length of a straight line connecting the start and end locations. Both lengths were collected directly by our tracking software.

3 Results

The results were processed using two-way ANOVA with the guidance system type (AR vs. US-only) and the trial number (1 to 4) as factors. To correct for multiple comparisons, the Holm-Bonferroni correction was used with a combined

significance of $p \leq 0.05$. The results displayed in Fig. 5 demonstrate that both the system type and trial number have a significant effect on both metrics with no significant interaction effect.

Box-plots organizing the data by the guidance system type are shown in Fig. 5. The 95 % confidence interval on the average improvement in terms of time for AR compared to US-only guidance is 3.51 ± 1.44 s. The 95 % confidence interval on the average improvement in terms of needle path tortuosity for AR compared to US-only guidance is 150 ± 40 %. Thus, AR guidance shows a significant improvement in terms of procedure time and needle path tortuosity. In addition, AR showed a lower interquartile range than US-only guidance for both metrics, implying that performance in AR was more consistent across participants.

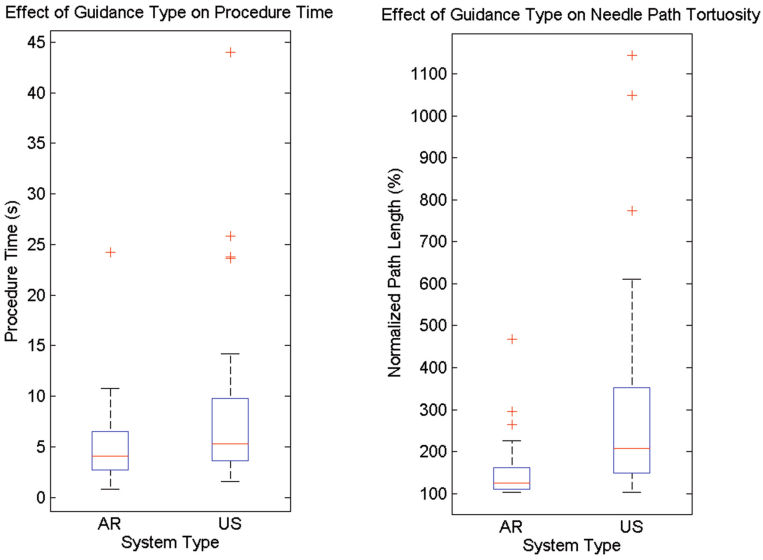


Fig. 5. Results organized by guidance system type. AR significantly outperforms US in terms of both time and tortuosity.

A significant training effect was detected in our analysis of variance. To examine this effect, the time and tortuosity results re-organized by trial are plotted in Fig. 6. Because there was no significant interaction between system type and trial number, Fig. 6 shows the results for ultrasound-only and AR guidance combined. For both metrics, there is a clear improvement over the first three trials which qualitatively confirms the existence of a strong training effect for novice users for the procedure as a whole. For both time and tortuosity, the interquartile range decreased over the first three trials while keeping a consistent lower bound indicating that not only was performance better on average, it is also more consistent amongst participants.

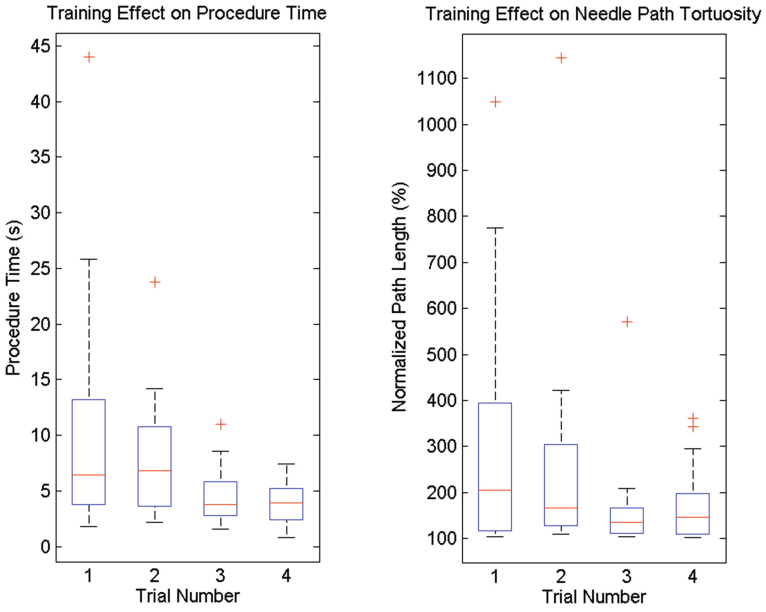


Fig. 6. Results organized by trial number. A significant training effect was observed for both metrics which is evident from the first three trials.

In terms of safety, there was only one incident in which the phantom CA was punctured, and this trial occurred under US-only guidance. A larger study is required to determine if there is a significance difference in terms of the number of carotid punctures given their infrequency under both guidance modalities.

4 Discussion

This study is a precursor to evaluating this system using medical practitioners and nursing staff who routinely perform central line cannulations. In this preliminary study, to establish the validation of the methodology, we investigated the performance of novice users guiding needle in a simulated central line procedure using an in-house designed phantom. The results indicate that AR guidance significantly outperforms traditional US-only guidance in terms of both time and needle path tortuosity, implying that the AR can make the central line procedure easier while minimizing patient discomfort. Additionally, AR showed more consistent performance that US-only guidance with fewer outliers and a tighter inter-quartile range, suggesting that the AR system can accommodate for variations amongst interventionalists, such as in the interpretation of the US images. Although these results are preliminary and use a novice participant group, they are promising as the AR guidance system enters clinical trials with expert interventionalists.

In addition to its use in an interventional setting, the proposed AR system could also be used for teaching purposes. As shown by Blaivas *et al.* [3], adequate training in central line procedures is a major concern deserving of greater attention. In this context, the AR system would teach the trainees how to interpret the ultrasound image with respect to the underlying anatomy and needle position so that they would be better able to perform the procedure using conventional US guidance. Such AR reality training has been shown to be beneficial for US-guided facet joint injections where it improved the safety and success rate compared with traditional US-only training [11]. While the main focus of our study was on the use of AR for guiding interventions, we did observe substantial training effects with the novice users. Our future work will better characterize these effects in central line procedures both with AR and US-only guidance and to repeat this experiment with expert interventionalists to better gauge clinical applicability.

5 Conclusion

In this paper, we propose and perform preliminary validation on an augmented reality (AR) guidance system tailored for the central line procedure. This AR system combined a magnetically tracked needle and ultrasound probe into a lightweight framework that is run directly from the ultrasound scanner itself. This system has been shown to significantly reduce procedure time and path tortuosity, improving the safety of the procedure. This system has received regulatory approval and is scheduled for clinical trials.

Acknowledgments. The authors would like to acknowledge John Moore for his invaluable assistance with the rapid prototyping of the support blocks for the phantom. The authors would also like to acknowledge the following funding agencies: Canadian Institutes of Health Research (CIHR), Canadian Foundation for Innovation (CFI), and Ontario Research Fund (ORF).

References

1. Ameri, G., McLeod, A., Baxter, J.S., Chen, E.C., Peters, T.M.: Line fiducial material and thickness considerations for ultrasound calibration. In: SPIE Medical Imaging, pp. 941529–941529. International Society for Optics and Photonics (2015)
2. Bannon, M.P., Heller, S.F., Rivera, M.: Anatomic considerations for central venous cannulation. Risk Manag. Healthc. Policy **4**, 27 (2011)
3. Blaivas, M., Adhikari, S.: An unseen danger: frequency of posterior vessel wall penetration by needles during attempts to place internal jugular vein central catheters using ultrasound guidance. Crit. Care Med. **37**(8), 2345–2349 (2009)
4. Denys, B.G., Uretsky, B.F.: Anatomical variations of internal jugular vein location: impact on central venous access. Crit. Care Med. **19**(12), 1516–1519 (1991)
5. Golden, L.R.: Incidence and management of large-bore introducer sheath puncture of the carotid artery. J. Cardiothorac. Vasc. Anesth. **9**(4), 425–428 (1995)

6. Krejza, J., Arkuszewski, M., Kasner, S.E., Weigele, J., Ustymowicz, A., Hurst, R.W., Cucchiara, B.L., Messe, S.R.: Carotid artery diameter in men and women and the relation to body and neck size. Stroke **37**(4), 1103–1105 (2006)
7. McGee, D.C., Gould, M.K.: Preventing complications of central venous catheterization. New Engl. J. Med. **348**(12), 1123–1133 (2003)
8. Odasso, D.P.: Internal jugular vein and carotid artery anatomic relation as determined by ultrasonography. Anesthesiology **85**(1.1111), 35–38 (1996)
9. Parsons, A.J., Alfa, J.: Carotid dissection: a complication of internal jugular vein cannulation with the use of ultrasound. Anesth. Analg. **109**(1), 135–136 (2009)
10. Tartière, D., Seguin, P., Juhel, C., Laviolle, B., Mallédant, Y.: Estimation of the diameter and cross-sectional area of the internal jugular veins in adult patients. Crit. Care **13**(6), R197 (2009)
11. Yeo, C.T., Ungi, T., Lasso, A., McGraw, R.C., Fichtinger, G., et al.: The effect of augmented reality training on percutaneous needle placement in spinal facet joint injections. IEEE Trans. Biomed. Eng. **58**(7), 2031–2037 (2011)

Interaction-Based Registration Correction for Improved Augmented Reality Overlay in Neurosurgery

Simon Drouin(✉), Marta Kersten-Oertel, and D. Louis Collins

Montreal Neurological Institute, Montréal, Canada
simon.drouin@mail.mcgill.ca

Abstract. In image-guided neurosurgery the patient is registered with the reference of a tracking system and preoperative data before sterile draping. Due to several factors extensively reported in the literature, the accuracy of this registration can be much deteriorated after the initial phases of the surgery. In this paper, we present a simple method that allows the surgeon to correct the initial registration by tracing corresponding features in the real and virtual parts of an augmented reality view of the surgical field using a tracked pointer. Results of a preliminary study on a phantom yielded a target registration error of 4.06 ± 0.91 mm, which is comparable to results for initial landmark registration reported in the literature.

1 Introduction and Background

In image-guided neurosurgery (IGNS), the head of the patient is rigidly attached to the operating room table and fixed relative to the frame of reference of a tracking system. The preoperative images can be registered with the patient's anatomy using different methods, including the selection of predefined anatomical landmarks on the patient using a tracked pointer and skin surface matching. This registration procedure is performed before sterile draping of the patient, since relevant features are not typically accessible after draping is completed. After this initial patient registration however, there can be a significant loss of navigation accuracy due in part to draping, attachment of skin retractors, and the duration of surgery as reported in [1]. In addition, registration accuracy is also affected by '*brain shift*', which can be caused by a number of factors including CSF drainage, swelling and resection. Brain shift at the cortical surface can range from almost no detectable shift up to 50 mm [2].

Several solutions have been proposed to improve patient-to-image registration during surgery, including: intraoperative MRI, intraoperative ultrasound with automatic registration to preoperative data [3] and computer-vision based techniques to register the surface of the operating field with preoperative data [4, 5].

In this paper, we propose a method that allows the surgeon to correct patient registration manually without having to remove his attention from the surgical field or the need to introduce additional equipment in the operating room (OR). To do so, we rely on a surgical microscope, the navigation system, and the tracked navigation pointer, which are already present in the OR in IGNS. The tracked surgical microscope

C.A. Linte et al. (Eds.): AE-CAI 2015, LNCS 9365, pp. 21–29, 2015.
DOI: 10.1007/978-3-319-24601-7_3

is used to produce an augmented reality (AR) image. It provides the surgeon with a single image that contains corresponding features from the patient and the preoperative data, allowing him to visualize the discrepancy. The surgeon uses the tracked navigation pointer to trace corresponding features in both images. These traces can then be used to establish a correction matrix for the patient registration.

The main contribution of this paper lies in its innovative use of AR to allow the surgeon to specify corresponding features on the patient and in preoperative data directly within his field of view without having to rely on the help of a technician. To our knowledge, it is the first time AR is used in this way to improve patient registration in IGNS.

The advantages of the proposed registration paradigm are threefold:

1. The surgeon can correct the registration at any moment during the surgery without having to remove his attention from the surgical field and without the intervention of a technician.
2. The method is robust because it is based the surgeon's extensive knowledge of the anatomy and of the specificities of the patient on the operating room table.
3. In the future, this method could be used to provide a starting point for automated methods that might further refine the patient-to-image registration.

In neurosurgery, the features that are most likely visible in both rendering of preoperative scans and live video of the operating field are sulci and blood vessels. Although the method presented in this paper can apply to both types of features, we focus our attention on blood vessels.

2 Materials and Methods

In this section we first describe the surgical context in which our AR-based registration method can be used. Then we give an overview of the system that is used to produce AR images before describing the registration method itself.

2.1 Surgical Context

Figure 1 illustrates the surgical context in which our method is used. The patient is rigidly attached to the operating table by way of a Mayfield® clamp for example, and a reference tool acts as the origin of the IGNS system's frame of reference. The patient's preoperative imaging data is registered to this coordinate system, typically using a patient-to-image landmark registration that yields transform P.

An AR view is obtained by merging live video images captured from the microscope (the real image) and a 3D volume rendering of preoperative patient data computed from the point of view of the microscope (the virtual image). Before rendering, the patient data is transformed to the space of the microscope's optics by concatenating following transforms: (1) P, the patient registration, (2) M, the microscope transform obtained directly from the tracking system and (3) E, the extrinsic calibration transform

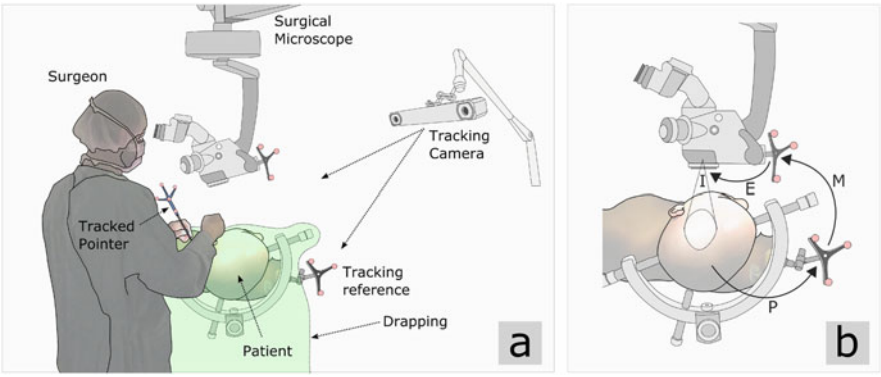


Fig. 1. (a) Surgeon using a tracked surgical pointer to trace features of the anatomy with the help of an AR view displayed within the microscope oculars or on the navigation system. The AR view is obtained by combining live video images from the tracked microscope and 3D rendering of preoperative images registered to the reference of the tracking system. (b) Transformation model used to render preoperative images from the point of view of the microscope: *P*: Initial patient to IGNS system registration transform, *M*: Transform between the IGNS system reference and the tracker tool attached to the microscope, *E*: Extrinsic calibration transform that maps the tracker tool to the optical center of the microscope, *I*: intrinsic calibration transform that projects 3D points in microscope space to the image plane.

discussed below. Once the data is in microscope space, it can be rendered using a standard direct volume rendering technique and a perspective projection model *I* whose parameters are estimated during a preoperative calibration procedure also described below.

2.2 Microscope Calibration

The preoperative microscope calibration procedure enables the estimation of the projection model of the microscope's optics *I* as well as the rigid transform *E* between the tracker tool attached to the microscope and the optical center of the microscope. The calibration procedure consists of capturing a series of microscope images of a checkerboard pattern printed on a flat board to which we rigidly attached a tracker tool compatible with the IGNS system. For every image, we record the transform of the tool attached to the board. The parameters of the optical model of the microscope are estimated using OpenCV's implementation of the method presented by Zhang [6]. The extrinsic transform *E* is obtained by combining microscope poses computed by Zhang's method and the tracker tool transforms recorded from the IGNS system. An optimization procedure borrowed from the field of robotics allows for the simultaneous computation of (1) the transform from the tracker tool to the board and (2) the extrinsic calibration matrix *E*. For more details about the optimization procedure, we refer the reader to [7].

2.3 Merging Real and Virtual Images

The process of merging real and virtual images to produce the final AR view is illustrated in Fig. 2.

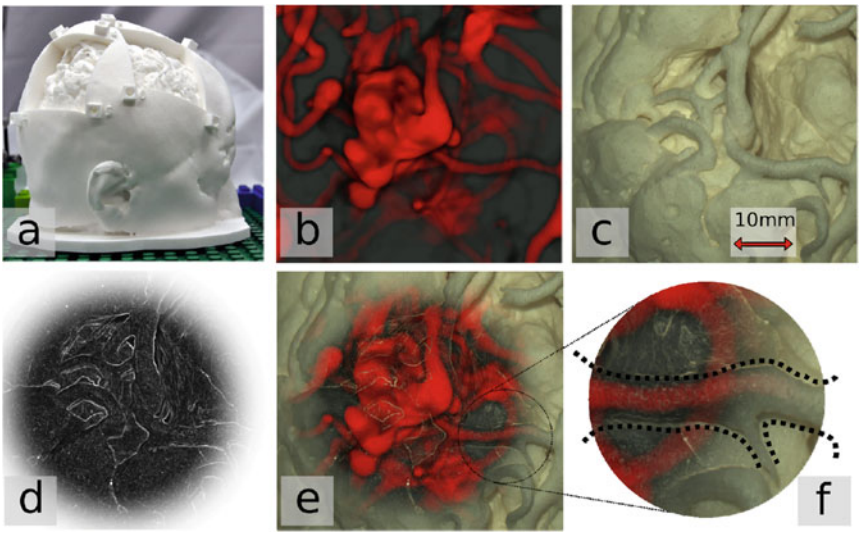


Fig. 2. (a) Phantom used to illustrate the method. (b) Virtual image rendered from the point of view of the microscope. (c) Real image captured from the microscope. (d) Mask that is used to determine the opacity of real image per pixel. (e) Resulting AR view obtained by combining the masked real image and the virtual image. (f) Close-up on a vessel that shows alignment of real and virtual images (diameter of circle is $\sim$ 12 mm).

The AR view in this example is produced with the 3D nylon printed patient phantom shown in Fig. 2a. Parameters of the tracked surgical microscope obtained by way of the calibration procedure outlined in the previous section are used to produce a 3D rendering of preoperative patient data from the point of view of the microscope (Fig. 2b). After capturing an image from a USB digital camera (FireFly MV, Pointgrey, Richmond, BC, Canada), attached to one of the optical ports of the microscope (Fig. 2c), we compute a mask (Fig. 2d) that is used to alpha-blend the real and virtual images to produce the final AR view (Fig. 2e). The mask is created by computing the pixel-wise maximum opacity between a blurred circular transparent region and Sobel-filtered version of the real image. The center of the circular region is updated in real-time to follow the projection of the tip of the tracked surgical pointer on the microscope image, allowing the surgeon to control the area of the real image that is transparent. The Sobel filter is used to extract edges in the real image to maintain occlusion cues and create the perception that the elements of the virtual image are located below the surface rather than floating above it, a problem often reported with augmented reality images [8].

2.4 Curve Tracing and Registration

Once we have produced an AR view of the surgical field in the OR, we can apply our method to correct for the misalignment between real and virtual images discussed above. The system allows the surgeon to use the tracked surgical pointer of the IGNS system to trace one or more corresponding piecewise linear curves in the real and virtual parts of the AR images as illustrated in Fig. 3.

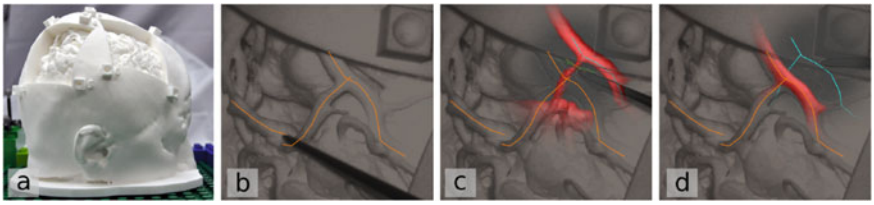


Fig. 3. (a) MR + CTA-based phantom with simulated craniotomies exposing the cortex and superficial blood vessels. (b) Using the tracked surgical pointer of the IGNS system, the surgeon can trace piecewise linear curves (orange curves) along the surface of the vessels. (c) The area around the surgical pointer becomes transparent, revealing corresponding misregistered vessels in the CTA, which can be traced in a similar way. (d) After both real and virtual images have been traced, the curves can be registered using the iterative closest point algorithm. Applying the resulting transform to the CTA aligns it with the virtual image (Color figure online).

Curves on the real image are traced by simply moving the surgical pointer along the surface of the tissues of interest and capturing the 3D position of the tip of the pointer. The surgeon triggers the acquisition of control points of the curve by pressing a USB foot pedal connected to the navigation system. The use of a foot pedal allows to avoid bringing a new piece of equipment within the sterile field.

Capturing the corresponding curve in the virtual image is slightly more complicated. Misregistration of the patient might cause the features of interest to lie below the surface of the patient's tissues for example. In this case, it is not possible to reach those areas with the tip of the pointer.

To determine the exact position of the point to capture, we use the concept of 3D picking. When the user presses the foot pedal, a ray is traced starting from origin of the virtual camera, going through the pointer tip and find the first vessel along the line of sight. The vessel is identified by finding the first voxel along the ray with intensity higher than a predefined threshold. The point that is picked on the vessel becomes the virtual coordinate of the next curve control point. This method allows tracing of elements of the virtual image without having to touch the tissues with the tip of the pointer.

Once the curves have been traced on the real and AR images, the corresponding curves are used to compute a correction of the initial patient registration using the iterative closest point (ICP) algorithm [9]. The registration transform computed is thus rigid. Furthermore, since control points in one curve are not matched to the closest control point in the other curve but rather to the closest location along the curve, the

number of points in both datasets don't need to match. In this work, we use the open source implementation of the ICP algorithm provided in the Visualization Toolkit (VTK) software package.

3 Experiment

We validate our method with a simple user study in the laboratory. The goal of the study is to show that registration accuracy can be improved with our method in a controlled lab environment. We test our method using a 3D printed phantom that is based on MRI and CT DSA imaging of a patient operated for the ablation of an AVM at the Montreal Neurological Hospital (Fig. 2a). The phantom represents the whole head of the patient and has simulated craniotomies that expose the cortex and superficial blood vessels. It was designed with 8 conical recesses around the simulated craniotomies that are used as landmarks. The position of the apex of the recesses is known, which allows for a very accurate landmark registration of the phantom with preoperative data. For more details about the fabrication of the phantom, we refer the reader to [10].

Prior to the experiment, we registered the phantom to its CT data by capturing the world space position of the phantom's built in landmarks with the tracked pointer. The registration transform is then computed using Horn's method [11]. We obtained a fiducial registration error (FRE) of 1.12 mm. The tracked microscope has also been calibrated according to the method described above. A cross-validation yielded a reprojection error of 0.37 mm for camera calibration. After completing these 2 initial steps, we are able to produce an accurate AR view of the simulated craniotomy of the phantom.

The user study consists in every subject attempting to correct simulated patient mis-registration 5 times using our method. The subject is initially trained and asked to explore the AR view to find vessels that are visible in both the real image and the rendered image. The subject is then asked to trace the surface of those vessels in the real image.

For each of the trials, we apply an artificial offset transform to the patient's preoperative data, simulating the loss of navigation accuracy that can result from initial phases of the surgery. The subject then needs to correct for this offset by tracing blood vessels on the virtual part of the image. The offset transform is composed of a translation and a rotation. The translation is obtained by choosing a random direction in the plane perpendicular to the optical axis of the microscope. The rotation is defined around the same axis and the sign of the angle is chosen randomly. The amplitude of the rotation and translation for each of the trial are listed in Table 1. One of the hypotheses we pose in this study is that our method may improve the registration only for shifts larger than a certain threshold. For this reason, the amplitude of the artificial shift we used in the experiment is decreasing with every trial. The maximum values for amplitudes are motivated by practical reasons. In the OR, shifts larger than 15 mm can happen and have been reported in the literature. However, random shifts of larger amplitude can cause the features of the virtual image to be out of the field of view of the microscope. If such case should happen in the OR, the surgeon could reposition the

microscope and would still be able to use our method. However, for the purpose of our analysis it was not possible to move the microscope during this experiment.

Table 1. Amplitude of the offset for translation and rotation of each of the trials of the study.

Trial no	Translation amplitude (mm)	Rotation amplitude (deg.)
1	15	5
2	10	4
3	6	3
4	3	2
5	1	1

4 Results

We ran the user study described above with 5 subjects who are all medical imaging experts. We used the set of 8 landmark points embedded in the phantom to measure the accuracy of the registration correction obtained with our method. We compute d_{off}, the distance between the original landmark position and its position after imposing the artificial offset (red cross in Fig. 4a), and d_{cor}, the distance between the original landmark position and their position after applying the correction computed using the proposed method (green cross in Fig. 4a). For each of the trials, we compute the root mean square (RMS) of d_{off} and d_{cor} over the 8 points. Figure 4b shows a plot of the resulting $RMS(d_{cor})$ as a function of $RMS(d_{off})$, where each of the points represents one trial of one of the subjects. This plot is an indication of how registration accuracy of our method varies with the original offset. We also computed the mean RMS corrected distance over all trials and all subjects and obtained 4.06 ± 0.91 mm.

5 Discussion

Results of this preliminary study show that the initial offset distance has little influence on the accuracy of the resulting registration after the proposed manual correction. This suggests that our technique could be used to correct for arbitrarily large misalignment of the patient with the preoperative data, such as when the navigation setup is accidentally displaced during the operation.

In [1], Stieglitz *et al.* reviewed the literature on accuracy of patient registration. They report errors ranging between 2.7 and 6.2 mm, with a median of 4.0 mm. The mean registration error obtained with our method (4.06 ± 0.91 mm) is thus comparable with the outcome of standard initial registration methods.

In this study, each subject was asked to perform the task only 5 times. A greater number of trials per subject would be desirable, but in practice, since we have only 1 phantom available, we found that subjects tend to produce the same trace for every trial and 5 trials per subject was sufficient to account for the variability of the traces that can be obtained. In a future study, we will perform each trial with a different phantom.

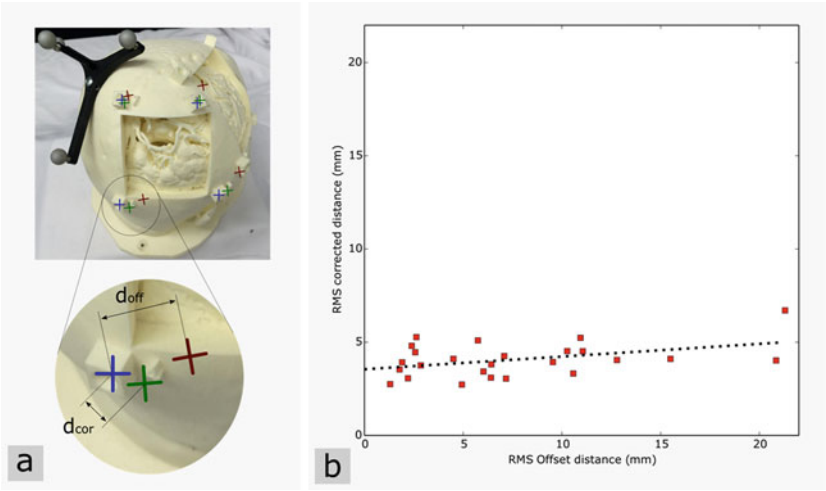


Fig. 4. (a) Illustration of the landmarks that are used to compute RMS offset distance and RMS corrected distance. Blue crosses show the position of the original landmarks, red crosses show the position of the landmarks after applying the artificial offset and the green crosses represent the position of the landmarks after applying our method. (b) Corrected RMS distance as a function of the RMS offset distance for each trial (red squares) and corresponding linear fit (black dashed line) (Color figure online).

One of the findings from our study is that the curves traced by the subjects are very noisy. This might be due to the relatively primitive tracing tools available so far in our system. If the tools are refined, by allowing for Bezier curves or by using computer vision methods to automatically snap the curve to features of images, it will be possible to significantly improve the accuracy of the registration.

6 Conclusion and Future Work

We have presented a simple system that can allow surgeons to correct the loss of navigation accuracy during an operation by leveraging their knowledge of the anatomy and taking advantage of a set of tools already present in the OR (tracked surgical pointer, surgical microscope, and navigation system). Through a user study in the lab, we have shown that our technique can produce registration accuracies comparable to state of the art methods used for initial registration of the patient before surgical draping. The next step is to bring our system to the OR where its accuracy could be compared to other registration correction methods.

One of the main advantages of our method is its robustness that comes from the fact that it relies on the surgeon's knowledge of the anatomy and it is inherently manual. In the future, we would like to study how this robust method can be used to constrain other more automatic methods such as ultrasound-based automatic registration. It would be particularly interesting to use the curves traced with our method to regularize

the computation of non-linear registration between preoperative MR scans and intra-operative ultrasound and correct for brain shift.

References

1. Stieglitz, L.H., Fichtner, J., Andres, R., Schucht, P., Krähenbühl, A.-K., Raabe, A., Beck, J.: The silent loss of neuronavigation accuracy: a systematic retrospective analysis of factors influencing the mismatch of frameless stereotactic systems in cranial neurosurgery. Neurosurgery **72**, 796–807 (2013)
2. Nabavi, A., Black, P.M., Gering, D.T., Westin, C.F., Mehta, V., Pergolizzi, R.S., Ferrant, M., Warfield, S.K., Hata, N., Schwartz, R.B., Wells III, W.M., Kikinis, R., Jolesz, F.A.: Serial Intraoperative MR Imaging of Brain Shift. Neurosurgery **48**, 787–798 (2001)
3. Reinertsen, I., Lindseth, F., Askeland, C., Iversen, D.H., Unsgård, G.: Intra-operative correction of brain-shift. Acta Neurochir. (Wien) **156**, 1301–1310 (2014)
4. Berkels, B., Cabrilo, I., Haller, S., Rumpf, M., Schaller, K.: Co-registration of intra-operative brain surface photographs and pre-operative MR images. Int. J. Comput. Assist. Radiol. Surg. **9**, 387–400 (2014)
5. Ji, S., Fan, X., Roberts, D.W., Hartov, A., Paulsen, K.D.: Cortical surface shift estimation using stereovision and optical flow motion tracking via projection image registration. Med. Image Anal. **18**, 1169–1183 (2014)
6. Zhang, Z., Member, S.: A flexible new technique for camera calibration. IEEE Trans. Pattern Anal. Mach. Intell. **22**, 1330–1334 (2000)
7. Dornaika, F., Horaud, R.: Simultaneous robot-world and hand-eye calibration. IEEE Trans. Robot. Autom. **14**, 617–622 (1998)
8. Lerotic, M., Chung, A.J., Mylonas, G., Yang, G.-Z.: Pq-space based non-photorealistic rendering for augmented reality. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, vol. 10, pp. 102–109 (2007)
9. Zhang, Z.: Iterative point matching for registration of free-form curves (1992)
10. Drouin, S., Kersten-Oertel, M., Chen, S.J.-S., Collins, D.: A realistic test and development environment for mixed reality in neurosurgery. In: Linte, C.A., Moore, J.T., Chen, E.C., Holmes III, D.R. (eds.) AE-CAI 2011. LNCS, vol. 7264, pp. 13–23. Springer, Heidelberg (2012)
11. Horn, B.K.P.: Closed-form solution of absolute orientation using unit quaternions. J. Opt. Soc. Am. A: **4**, 629 (1987)

'*On the Fly*' Reconstruction and Tracking System for Patient Setup in Radiation Therapy

Hagen Kaiser(✉), Pascal Fallavollita, and Nassir Navab

Technische Universität München, Munich, Germany
hagen.kaiser@brainlab.com

Abstract. Range imaging devices have already shown their value for patient setup and motion management in external beam radiation therapy. However current systems need several range imaging devices recording the patient's surface from different viewpoints to achieve the required stability and accuracy. Since range imaging devices come as add-ons to regular linear accelerators, they have to share the limited space with other sensors in the treatment room to get a line of sight to the patient's isocenter. The objective of this work is to describe a new registration framework which enables stable tracking using only one range imager. We unveil the design of our solution to the problem of tracking a patient over long trajectories and large viewpoint changes including surface acquisition, pose estimation and simultaneous surface reconstruction. We evaluate the performance of the system using three clinically motivated experiments: (i) motion management, (ii) non-coplanar patient setup and (iii) tracking over very large angles. We compare our framework to the state-of-art ICP algorithm and to a ground-truth stereoscopic X-ray system from BrainLab. Results demonstrate that we could track subtle movements up to 2.5 cm with a mean target registration error of 0.44 mm and 0.02°. Subsequent non-coplanar field setup on 30° and 2 cm motion yielded a target registration error of 2.88 mm which is within clinical tolerances. Our sensor design demonstrates the potential of simultaneous reconstruction and tracking algorithms and its use for patient setup and motion management in radiation therapy.

Keywords: Range imaging · Patient setup · Motion management · Radiation therapy

1 Introduction

Surface scanning systems that utilize a pre-op CT surface for patient setup in radiation therapy can cause target registration errors up to 1.3 cm [8]. A typical clinical protocol to circumvent this is to combine CBCT with surface imaging. In this scenario one CBCT is used to setup the patient. Immediately after CBCT acquisition a surface imaging device records the position of the patient as a reference surface. From that instant, patient movements can be tracked and compensated. In the event that the patient has to be treated at multiple anatomical sites or additional non-coplanar beams, the range imager can also be used for subsequent site setup without further CBCT acquisition. To realize that, a range imaging tracker must be able to detect subtle movements of a patient but also track the patient over large trajectories and rotations.

C.A. Linte et al. (Eds.): AE-CAI 2015, LNCS 9365, pp. 30–37, 2015.
DOI: 10.1007/978-3-319-24601-7_4

Using a single sensor, large areas within an initially acquired reference surface become invisible while the relative viewpoint changes. This would cause a state-of-art ICP registration algorithm to fail.

1.1 Literature Review

A small number of optical, markerless systems have been developed over the last years, using different physical methods such as Kinect [1], speckle projection photogrammetry [2], time-of-flight (ToF) measurements [3, 4] and laser triangulation [5]. Most of these techniques register the reconstructed surface to pre-op CT for patient positioning. Authors using the Kinect sensor for patient setup focused only on bringing the patient roughly from the loading position, where the patient is supposed to enter the couch, to the isocenter position. The translational errors of these approaches are 1 cm. In 2013, a positioning system was developed, consisting of two camera and projector units capable of projecting and capturing 120 images/s and covering 180° of the surface [6]. On the commercial side, the stereo system AlignRT (*VisionRT, London UK*) uses three camera pods to cover as much as possible of the patients surface at once. The system was reported in [7] to have a mean translational registration error of 1.2 mm. It needs 3–10 s for patient position estimation, which is too slow to account for spontaneous movements. Recently, the Catalyst (*C-RAD, Uppsala Sweden*) system was released containing multiple cameras-rigs for full body coverage. It however needs three devices to be mounted, which can often block the field of view of other systems in the treatment room. The system was studied in [8] and reported a mean accuracy of 4 mm and 1.7°. To our knowledge, all clinically applied systems today, that are used for 'real-time tracking' of a patient, merge range images from several sensors at different viewpoints to get sufficient coverage of the patient's surface during radiation therapy.

1.2 Contributions

While state-of-art systems above stabilize patient positioning and tracking by stacking up the hardware and merge the images of several range imaging devices, we solve that problem algorithmically needing only one sensor.

We emphasize several advantages of the proposed system: (**1**) the cost of our hardware < 1000 k compared to commercial systems as VISION RT ($\sim$ 250,000 k); (**2**) compared to '*one stereo-cameras*' sold by BrainLAB that are marker-based requiring patient preparation, our system is seamless with no patient preparation; (**3**) other commercial markerless stereo trackers come as multi-camera setups whereas we propose a single camera rig; (**4**) the accuracy is tested on both planar and non-coplanar experiments – a first within the MICCAI community; (**5**) our core contribution is how we use the patient generated point clouds. We iteratively extend the patient model as more of the patient's body becomes visible. In this way we increase the body coverage; and (**6**) the runtime varies with how many points are in the patient model. Our system performs with a higher frame rate on average, outperforming AlignRT (0.1–0.3 fps), a commercial system. To validate the performance and accuracy, an in-depth evaluation

is presented on 6D pose recovery of our system compared to a ground-truth stereoscopic X-ray technology from BrainLab.

2 Methods

From Fig. 1-top, our patient positioning and tracking pipeline is separated into three threads: (a) stereo reconstruction, (b) tracking, and (c) parallel patient model reconstruction. These three tasks can be performed independently as they have different demands regarding performance respectively. The tracking task is developed in real-time. We optimize a global patient model where we accumulate the information collected by the surface acquisition to get a more reliable reference surface constantly over time, which we use to align to the current measurement.

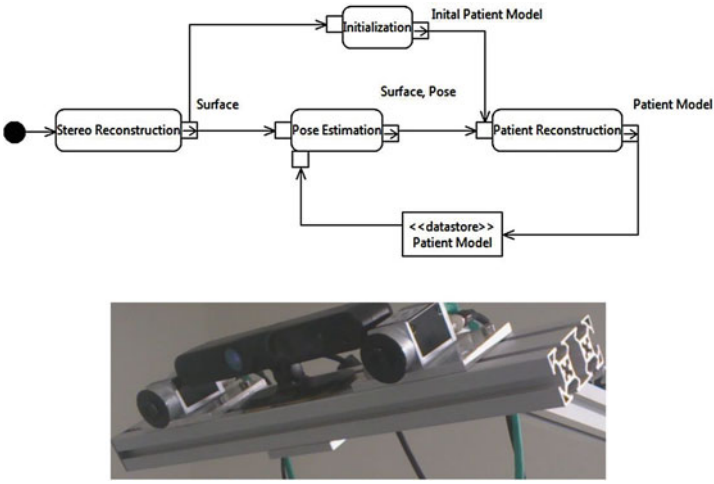


Fig. 1. (*Top*) Algorithm pipeline. (*Bottom*) The single stereo rig.

2.1 Surface Acquisition

As a range imaging device we use a stereo-rig made of two Basler cameras with a resolution of 1604 × 1204 pixels (Fig. 1-bottom). Each camera lens is augmented with an IR filter. The cameras are mounted with a baseline of 12 cm. To this setup, we add an IR laser with a fixed diffraction grating as available in the Kinect sensor. It projects a speckle pattern on the patient surface. The pattern is invisible to the patient and thus will not cause irritation unlike the pattern used by the system described in [7]. We mount the stereo-rig to the ceiling near the LINAC gantry at 1 m distance to the machine isocenter. This is an optimal position regarding compatibility with other scanning systems that are usually also mounted in the treatment room and need a line of sight to the machine isocenter. At time $\boldsymbol{k}$ the stereo rig, consisting of camera **a** and

camera **b**, acquires simultaneously two images from the projected pattern. To calculate the range image RI_k which is a set of point measurements $p_{u,v} \in \mathbb{R}^3$ indexed by the originating pixel (u, v) from camera **a**, we used the implementation of the census transform algorithm [9]. The algorithm has a low memory footprint, is fast, and yields satisfying accuracy. Hence, it is suited to be implemented on embedded platforms, which would reduce the size of the device even more. To match a pixel (u,v) from camera **a** to camera **b**, the algorithm searches along the epipolar line minimizing a cost function. A local neighborhood around pixel (u,v) is transformed into bitstring according to its intensity differences with

$$T(u,v) := \otimes_{i=-\frac{n}{2}}^{\frac{n}{2}} \otimes_{j=-\frac{m}{2}}^{\frac{m}{2}} \xi(I(u,v), I(u+i, v+j))$$

where I(u,v) is the intensity of (u,v), $\otimes$ is a bitwise catenation, (n × m) denoted the local neighborhood. The parameter ξ is defined as

$$\xi(x,y) := \begin{cases} 0 \; if \; x \leq y \\ 1 \; if \; x \geq y \end{cases}$$

As our cameras are mounted in a horizontal line we can minimize for the disparity value **d** with the following cost function

$$C(u, v, d) := \mathrm{Ham}(T_a(u, v), T_b(u + d, v))$$

where *Ham* is the Hamming distance between bitstrings T_a and T_b of the images from camera **a** and camera **b**. Using the disparity **d** we get for each pixel (u,v) a metric point measurement $p_{u,v}$ = (x,y,z) in $\mathbb{R}^3$.

2.2 Patient Pose Estimation

The incoming range image RI is filtered using a bilateral filter on the current depth-frame to remove noise while still preserving edges. We down sample the cloud further by discretizing the space into regular voxels of 0.3 mm^3 and taking only the average of each occupied voxel. As soon as the buffer that holds the current patient model has been filled, the tracking thread starts registering the model against the prepared range image RI. Therefore, we use a variant of the ICP algorithm initialized with the most recent pose by incorporating the plane-to-plane metric [10]. Using the plane-to-plane metric we are able to successfully exclude sliding effects which are introduced by the thorax of patients, for example, being a larger low-frequency region in the point cloud.

2.3 Patient Reconstruction

As we do not rely on a pre-operative CT surface, we have to acquire a 'close enough' initial guess of the surface in the initialization phase. We take advantage that following

the clinical procedure (i) the patient has been positioned at the isocenter and in the field of view of the camera rig and (ii) the patient is not moved in the initialization phase. Thus we can define a filter which rejects all points not being in a box sized region of 50 cm^3 around isocenter. Subsequently, we acquire 5 frames and feed them directly into the patient model reconstruction thread. We assume for all frames captured in the initialization phase the registered pose to be the origin pose. Later we qualify the registration result by their residual registration error and integrate the cloud if the error is reasonably low. The incoming frame is then segmented by a simple outlier rejector that removes points having less than 20 neighbors in a radius of 2 cm. We rely on the assumption that the patient is a single large cluster and other small clusters are noise and do not belong to the patient. The remaining points are assumed to belong to the patient only and will be integrated into a single global cluster cloud representation which we can query for either a complete version or a geometrically stable sample of the cloud [11]. Having performed the initial segmentation and integrating only points that are very near to our patient model, we can skip the segmentation step from now on. Each point of the reference surface serves as a seed for a corresponding cluster. The cluster cloud is a set $\mathbf{C}$ of clusters $\mathbf{Cl_i}$ with i = 1, 2,…n. Each cluster has a cluster representant $\mathbf{c_i} \in \mathbb{R}^3$. Furthermore, there exists a variance σ, and a generation $\mathbf{g_i}$ for each cluster, whereas the generation defines the number this cluster has been confirmed by subsequent observations. Querying the model will either yield the complete or a downsampled set of cluster representants from the cluster cloud. For each incoming registered frame we perform a Neaserst-Neighbour-Search on the complete version of the patient model between points $\mathbf{p} \in \mathbb{R}^3$ from the incoming frame and cluster representants $\mathbf{c}$. We threshold correspondences to a maximum distance value. This shall enforce that only points correspond to each other that represent a measure of the same point in space. Each cluster $\mathbf{Cl_i}$ which has a representant $\mathbf{c_i}$ that has corresponding nearest point $\mathbf{p_j}$ will be updated calculating the cumulative moving average. We then update the variance of the cluster using the sum of squared differences from the current representant $\mathbf{c_i}$. This can easily be incrementally updated and yields the variance of the cluster in a numerically stable way. Similar to [12] we use a data structure to fuse repeating measurements of the same scenery to improve the performance of the system. Using the cluster cloud we do not discretize space artificially. We furthermore have a representation that immediately can be used for tracking and thus can save the step of generating a reference cloud by raytracing a truncated signed distance function representation. On the other hand we acquire confidence values on our measurements and thus can further remove noise from the reference cloud. If we query now a reference that is supposed to be used by the tracking thread, the resulting set of cluster representants can be thresholded by variance and generation. We rely here on the assumption that representants that have been measured by several frames with low variance have a higher confidence. The downsampled version of the cluster cloud is finally sent to a buffer where it can be consumed by the tracking thread. Figure 2 depicts the visual pipeline of our reconstruction and tracking of a torso phantom.

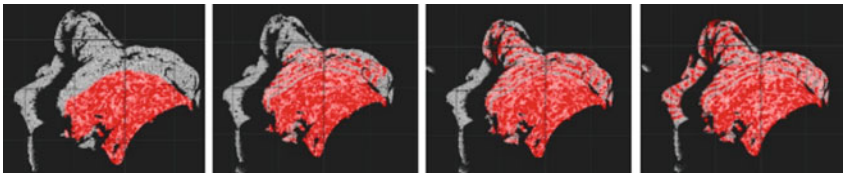


Fig. 2. The principle of surface evolution. Algorithm was initialized with torso only. While the patient moved along the longitudinal axis the patient was tracked in real-time and simultaneously reconstructed. (Grey color: Current Observation, Red color: Current patient model) (Color figure online)

3 Results and Discussion

Study Aim: Our aim was investigating the variability in patient movements that can happen during treatment (coplanar/non-coplanar), which is the primary concern in current clinical practice. Here, we define 'coplanar movements' as the patient being positioned for a coplanar beam (couch angle = 0°). The radiation therapy setting will never be utterly cluttered, and varying patient body size only implies varying size of the points in our model. Thus, we believe tuning the proposed system towards an anthropomorphic phantom is sufficient. Thus, we used a torso phantom which contains internally anthropomorphic bone structures. Lastly, we emphasize that aim to study the dynamic behavior of our system (not registration).

Ground-Truth: We evaluated our technology via a comparative study in a clinical setting against another clinical system (*Brainlab ExacTrac*, which is stereoscopic X-Ray). A CT scan was recorded and a test plan was created using *Brainlab IPlan RT* 4.5. The surface was filmed using our camera setup, while the internal structure was imaged and registered against the CT scan using the stereoscopic X-Ray, which became our ground truth. We positioned the torso phantom using ExacTrac in the machine isocenter. The target registration error is evaluated separately for the translational part TR_{trans} and the rotational part TRE_{rot}, where

$$TRE_{trans} = ||t_{alg}|| - ||t_{ref}||$$

t_{alg} is the translation computed by our algorithm and t_{ref} is the translation vector computed by the ground-truth stereoscopic X-Ray (similarly for the rotational part).

Study 1- Motion Management: Patient motion management implies the experienced subtle movements after the patient has already been correctly positioned. This largely happens when tension is relieved and the patient calms down prior to external beam radiation therapy. To simulate this, we initialized our algorithm to acquire the basis reference surface from which the point evolution can start. In the next step we displaced the torso phantom at twelve defined couch positions between 1 cm, 2 cm and 2.5 cm using an *Elekta Hexapod*. We verified the length of the offset using the *ExacTrac system* and compared these ground truth results to those of our algorithm (Table 1). The mean target registration error for the compensation of coplanar patient misalignment was 0.44 ± 0.1 mm. To evaluate the rotational component we rotated the

couch around the x, y, z axis for 2.5° respectively. The mean errors were: x = 0.049 ± 0.06°, y = 0.024 ± 0.03° and z = 0.032 ± 0.039°.

Table 1. Target registration errors for patient misalignment compared to ground-truth.

Shift (mm)	Lateral (mm)	Longitudinal (mm)	Vertical (mm)
10	0.077 ± 0.09	0.322 ± 0.08	0.527 ± 0.04
20	0.047 ± 0.16	0.389 ± 0.09	0.630 ± 0.05
25	0.160 ± 0.12	0.220 ± 0.17	0.748 ± 0.08

Study 2- Noncoplanar Patient Setup: In this experiment we investigate the common clinical scenario in which a patient has been setup using CBCT. The patient is prepared for a beam to the same treatment site but with a different couch angle. Hence, a solution must be in place to ensure the patient does not move. Furthermore, if the couch is not rotating perfectly at the isocenter then non-coplanar errors occur. To evaluate the system's performance with respect to non-coplanar positioning, we displaced the torso phantom manually from 0° to 30° around the couch's vertical axis while simultaneously tracking the phantom. We then repeated a test similar to the motion management of Study 1, but now moving the phantom for 2 cm along the vertical axis. Our algorithm must be robust to this accumulated drift. For an absolute shift of 30° vertical and 2 cm vertical displacement the algorithm performed with an error of 4.88 ± 0.721 mm. We compare this to a state-of-art ICP only registration, and observed that the ICP registration fluctuated already at 1.1 cm and a couch rotation of only 13°.

Study 3: Tracking Stability at Very Large Angles: We verified as well the stability of our system on very large angles. For this we tracked the phantom for 100° without

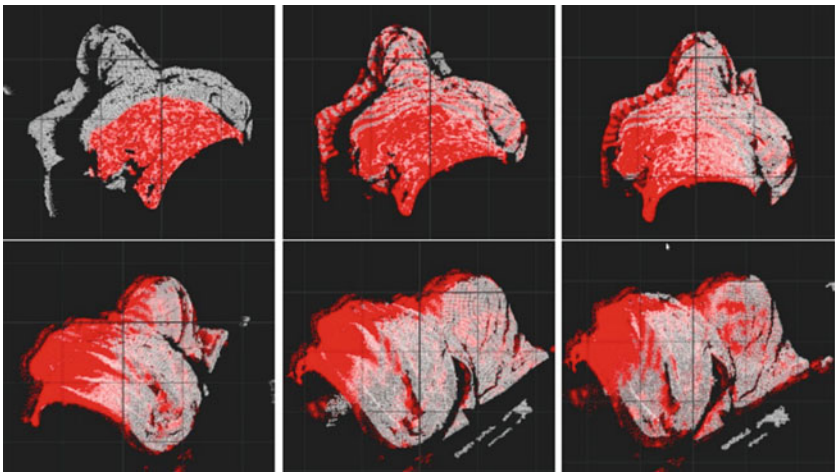


Fig. 3. Surface evolution during 100°couch angles. The grey points are the points currently detected by the sensor. The red points are distributed all over the surface although much of the surface in the final pose was initially not visible. (Grey color: Current Observation, Red color: Current patient model) (Color figure online)

falling into a local minimum (Fig. 3). With 80 % of the finally reconstructed points initially not visible, the algorithm tracked from 20 k to 100 k points with a speed of 2 Hz. As a comparison, the ICP-only registration diverged completely with couch angles larger than 57°.

4 Conclusions

We presented a new registration framework for patient setup and motion management in Radiation Therapy. We have demonstrated an algorithmic solution that can operate under limited fields of view using only a single camera compared to state-of-art technologies. The sensor consists of standard components and utilizes simultaneous reconstruction and tracking to achieve accuracy within clinical tolerances.

References

1. Bauer, S., et al.: Multi-modal surface registration for markerless initial patient setup in radiation therapy using microsoft's Kinect sensor. In: 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops). IEEE (2011)
2. Bert, C., et al.: Clinical experience with a 3D surface patient setup system for alignment of partial-breast irradiation patients. Int. J. Radiat. Oncol. Biol. Phys. **64**(4), 1265–1274 (2006)
3. Schaller, C., et al.: Time-of-flight sensor for patient positioning. In: SPIE Medical Imaging, pp. 726110–726110 (2009)
4. Placht, S., et al.: Fast time-of-flight camera based surface registration for radiotherapy patient positioning. Med. Phys. **39**(1), 4–17 (2012)
5. Brahme, A., et al.: 4D laser camera for accurate patient positioning, collision avoidance, image fusion and adaptive approaches during diagnostic and therapeutic procedures. Med. Phys. **35**(5), 1670–1681 (2008)
6. Lindl, B.L., et al.: TOPOS: A new topometric patient positioning and tracking system for radiation therapy based on structured white light. Med. Phys. **40**(4), 042701 (2013)
7. Peng, J.L., et al.: Characterization of a real-time surface image-guided stereotactic positioning system. Med. Phys. **37**(10), 5421–5433 (2010)
8. Stieler, F., et al.: A novel surface imaging system for patient positioning and surveillance during radiotherapy: a phantom study and clinical evaluation. Strahlenther. Onkol. **189**(11), 938–944 (2013)
9. Humenberger, M., et al.: A fast stereo matching algorithm suitable for embedded real-time systems. Comput. Vis. Image Underst. **114**(11), 1180–1202 (2010)
10. Segal, A., et al.: Generalized-ICP. Robot. Sci. Syst. **2**(4) (2009)
11. Gelfand, N., et al.: Geometrically stable sampling for the ICP algorithm. In: Proceedings of the Fourth International Conference on 3-D Digital Imaging and Modeling. IEEE (2003)
12. Newcombe, R.A., et al.: KinectFusion: real-time dense surface mapping and tracking. In: Mixed and Augmented Reality (ISMAR), pp. 127–136 (2011)

3D Catheter Tip Tracking in 2D X-Ray Image Sequences Using a Hidden Markov Model and 3D Rotational Angiography

Pierre Ambrosini[1(✉)], Ihor Smal[1], Daniel Ruijters[2], Wiro J. Niessen[1,4], Adriaan Moelker[3], and Theo van Walsum[1]

[1] Biomedical Imaging Group Rotterdam, Department of Radiology and Medical Informatics, Erasmus MC, Rotterdam, The Netherlands
p.ambrosini@erasmusmc.nl
[2] Philips Healthcare, Interventional X-ray Innovation, Best, The Netherlands
[3] Department of Radiology, Erasmus MC, Rotterdam, The Netherlands
[4] Imaging Science and Technology, Faculty of Applied Sciences, Delft University of Technology, Delft, The Netherlands

Abstract. Integration of pre- or peri-operative images may improve image guidance in minimally invasive interventions. In abdominal catheterization procedures such as transcatheter arterial chemoembolization, 3D pre-/peri-operative images contain relevant information, such as complete 3D vasculature, that is not directly available from 2D imaging. Accurate knowledge of the catheter tip position in 3D is currently not available, and after registration of 3D information to 2D images (angiographies), the registration is invalidated by breathing motion and thus requires continuous updates. We propose a hidden Markov model based method to track the 3D catheter position, using 2D fluoroscopic image sequences and a 3D vessel tree obtained from 3D Rotational Angiography. Such a tracking facilitates display of the catheter in the 3D anatomy, and it enables to use the 3D vessels as a roadmap in 2D imaging. The tracking is initialized with the first 2D image of the sequence. For the subsequent images, based on a state transition probability distribution and the registration observations, the catheter tip position is tracked in the 3D vessel tree using registrations to the 2D fluoroscopic images. The method is evaluated on simulated data and two clinical sequences. In the simulations, we obtain a median tip position accuracies up to 2.9 mm. On clinical sequence, the distance between the catheter and the projected vessels after registration is below 1.9 mm.

Keywords: Catheter · Tip · Tracking · Rigid · Registration · Guidance · X-ray · Fluoroscopy · 3DRA · Hidden markov model · Abdominal · TACE · Liver · Breathing

Electronic supplementary material The online version of this chapter (doi:10.1007/978-3-319-24601-7_5) contains supplementary material, which is available to authorized users. Videos can also be accessed at http://www.springerimages.com/videos/978-3-319-24600-0.

C.A. Linte et al. (Eds.): AE-CAI 2015, LNCS 9365, pp. 38–49, 2015.
DOI: 10.1007/978-3-319-24601-7_5

1 Introduction

Minimally invasive procedures are commonly performed to treat various diseases because they are less demanding and risky for the patient than, for example, open surgeries. In such procedures, as direct eyesight is lacking, physicians require intra-operative images to visualize the instruments and the anatomy. During catheterization interventions, 2D fluoroscopic (X-ray) imaging is usually used, but the noise, 2D projection and inability to continuously use contrast agent prevent physicians to have a continuous understanding of the instrument position with regard to the 3D vasculature.

The purpose of our work is to improve image guidance in 2D X-ray guided sequence abdominal catheterization procedures, specifically during transcatheter arterial chemoembolization (TACE) procedures, by using 3D information from peri/pre-operative images. During the procedure, the physician injects chemotherapeutic agents and embolizes liver tumors by inserting a catheter into the femoral artery and guiding it toward the tumors. The physician uses single-plane 2D X-ray images in which only the catheter and the ribs are visible (Fig. 1). Pre-operative 3D Computed Tomography Angiography (CTA) and intra-operative 2D angiographies (X-rays with contrast agent) are acquired, providing detailed images of the arterial tree, and enabling a roadmap to guide the catheter. However, such static roadmaps are hampered by breathing motion and catheter deformation. We therefore propose a 3D tracking method that follows the position of the catheter tip in the 3D vessel tree, enabling guidance in the 3D image as well as facilitating continuous roadmapping.

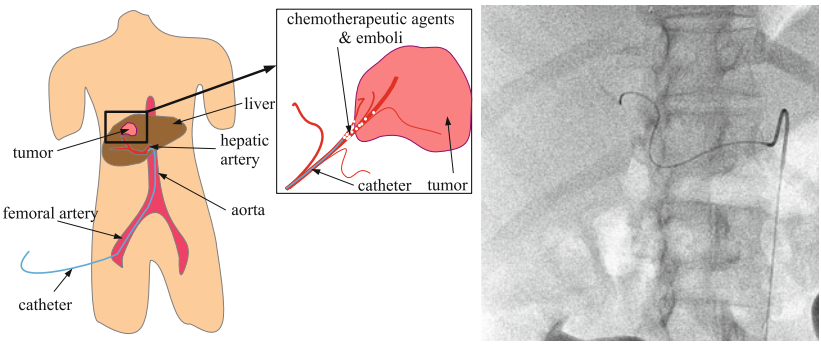


Fig. 1. TACE overview (left) and fluoroscopy example (right).

Image fusion and 3D/2D registration have already been addressed in the literature (see reviews [5,7]), particularly for 2D X-ray guidance in cranial [8], cardiac [10] and abdominal [3] interventions. These methods rely on anatomical structures such as bones or the vasculature. A vessel-based rigid or non-rigid registration is done between pre-/peri-operative 3D images and single/bi-plane 2D angiographies or 2D fluoroscopies. These approaches align the 3D vessels with the

2D vasculature visualized using contrast agent. Such approaches cannot be used continuously because of the toxic nature of the contrast agent. Without contrast agent, in cardiac interventions, Ma et al. [6] use features such as diaphragm/heart border, tracheal bifurcation or the catheter to correct for breathing motion. In 2D abdominal fluoroscopies, however, such features are lacking. Other methods propose 3D/2D registration with peri-operative 3D Rotational Angiographies (3DRA) or Cone Beam Computed Tomography (CBCT) [1,2,11] where the calibrated geometry of the C-arm enables accurate alignment with the 2D X-ray images (acquired with the same device). Such an approach is effective in cranial interventions, as the head does not deform [11], but in abdominal procedures, the breathing motion invalidates the alignment. That is why a semi-automatic method following a region of the catheter [2] and a catheter-based registration [1] have been proposed to follow the catheter. Ambrosini et al. [1] have an automatic registration but when the catheter visible part is too short, the alignment fails. Position tracking of catheters has been addressed less frequently. The method proposed by van Walsum et al. [13] is one of the first approaches for neural applications.

In our work, we combine a hidden Markov model (HMM) [9] with 3D/2D registration to track the catheter tip in 3D over the time. The main contribution is a novel method for tracking the catheter tip in 3D, using a 3D vessel tree, 2D images and a HMM. The method is evaluated using a large set of simulated catheter motions in patient 3D datasets, and demonstrated on two patient cases.

2 Method

The purpose of our method is to track the catheter tip inside a 3D vessel tree, where the catheter position in 2D is obtained from the interventional X-ray images. As the catheter is assumed to be in the vasculature tree and as its displacement is relatively small between subsequent 2D images, we propose to model the catheter motion within the 3D vessel tree using hidden Markov model (HMM) [9]. Each 3D point of the vessel centerlines represents a state that denotes the probability that the catheter tip is at that location. Each state is linked with state transitions between connected close-by vessel parts. The observations to update the HMM are based on a 3D/2D registration metric where the 3D vessel tree is aligned with the 2D catheter visible in the image.

In the following, we explain the HMM, followed by a description of how the different elements of the HMM are integrated in our 3D tracking method.

2.1 Hidden Markov Model

A HMM is described as a system with a set of states $S = \{s_1, \dots s_N\}$ (Fig. 2). The HMM changes, at each time point t, according to the probabilities associated with the states and the current set of observation $O_t = \{O_t(1), \dots O_t(N)\}$. The transition probabilities between states are defines in a matrix A (with a $N \times N$

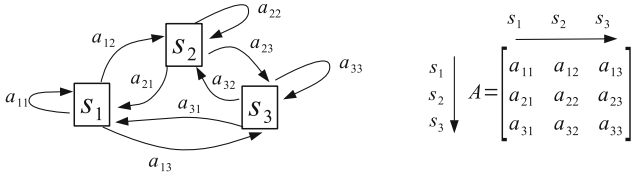


Fig. 2. HMM with 3 states and its matrix A of state transition probabilities.

dimension) where each $a_{ij} \in A$ is the probability that the state i can move to the state j ($a_{ij} \geq 0$ and $\sum_j a_{ij} = 1$).

Following Rabiner et al. [9], the Viterbi algorithm selects at time t the most probable path through the state space based on the maximum $\delta_t(i)$ which is the best score (highest probability) along a single path that ends in state s_i. Viterbi takes into account the first t observations. Starting from an initial distribution of the probabilities over the states $\pi = \{\pi_1, \ldots \pi_N\}$ where $\sum_{j=1}^{N} \pi_j = 1$, the algorithm initializes the $\delta_1(i)$ as follows:

$$\delta_1(i) = \pi_i \cdot O_1(i),$$

where $O_1(i)$ is an observation score given that we are in the state s_i at time $t = 1$. Next, the subsequent $\delta_t(i)$ can be computed using recursion:

$$\delta_t(j) = \max_i [\delta_{t-1}(i) \cdot a_{ij}] \cdot O_t(j).$$

2.2 Catheter Tip Tracking

Timepoint and States. At each timepoint t, a single 2D fluoroscopic image of the complete image sequence is processed. In this image, the 2D catheter centerline $C_t = \{c_1, \ldots c_{n_C}\}$ is extracted, where c_1 is the tip of the catheter. The full 2D catheter is used during the 3D/2D registrations to compute the observations. A set of 3D points $P = \{p_1, \ldots p_N\}$, corresponding to the 3D vessel centerlines, is extracted from the peri-operative 3DRA. In the HMM, the probability of being in state s_i is the probability of the catheter tip being at position p_i (Fig. 3).

Matrix A of State Transition Probabilities. Each a_{ij} in the matrix A contains the probability that the catheter tip moves from the position p_i to p_j (state s_i to s_j) between two images. In the context of tip motion, the closer p_i and p_j are, the higher the probability of transition should be.

To define A, the transition probabilities are set according to the distance along the vessel path between points p_i and p_j of the 3D vessel tree and distributed with a Gaussian function:

$$a'_{ij} = e^{-\frac{D(p_i, p_j)^2}{2\sigma_a^2}}$$

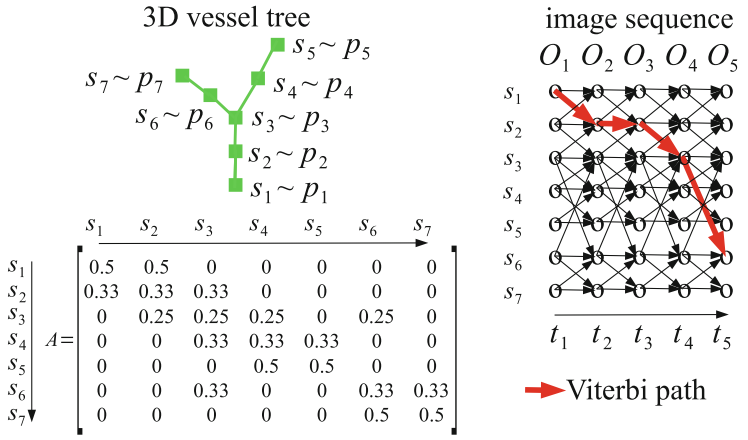


Fig. 3. HMM with a simple vessel tree. Here, the transitions between possible tip locations exist only with direct neighbours and are equiprobable. The Viterbi path goes to the optimum tip position knowing the observations $O_1 \dots O_5$ in the 5-images sequence.

where σ_a controls how far the catheter tip can move. If $\{l_1, \dots l_n\}$ is the set of points representing the vessel centerline between the points $p_i = l_1$ and $p_j = l_n$, $D(p_i, p_j)$ is defined as the sum of the distances between each neighboring pair l_k, and l_{k+1}:

$$D(p_i, p_j) = \sum_{k=1}^{n-1} ||l_k, l_{k+1}||.$$

Because the matrix A defines probabilities, $\sum_j a_{ij}$ has to be equal to 1 so we normalize the coefficients to obtain $a_{ij} = a'_{ij}.(\sum_j a'_{ij})^{-1}$.

Observations. An observation score $O_t(i)$, under the assumption that the catheter tip is at the position p_i (thus in the state s_i), has to be determined. From the 3D vessel tree, the unique 3D catheter path centerline $V_i = \{v_1, \dots v_{nV}\}$, starting from the tip p_i and going to the root of the tree, is extracted. This 3D path V_i is registered to the current 2D catheter C_t, obtaining a rigid transformation that aligns the 3D vessel tree with the catheter centerline in the 2D image. The observation score $O_t(i)$ (between 0 and 1) using the metric M of the 3D/2D registration is defined as follows:

$$O_t(i) = e^{-\frac{M(C_t, V_i)^2}{2\sigma_s^2}}$$

where σ_s controls the registration score distribution and where M is defined as the minimum sum of the minimal distance between each point of the 2D catheter

C_t and any projected point of the 3D centerline V_i:

$$M(C_t, V_i) = \min_{\tau} \left(\sum_{c \in C_t} \min_{v \in V_i} ||c - F_{\text{proj}}(v, \tau)|| \right)$$

where $F_{\text{proj}}(v, \tau)$ is the projection of the 3D point v onto the 2D images and τ represents a 4 degrees of freedom rigid transformation matrix (three rotations and one translation) used to align the catheter C_t and the vessel path V_i. We define the projection function F_{proj} as follows (Fig. 4):

$$F_{\text{proj}}(v, \tau) = T_{\text{proj}}.T_{\text{det} \leftarrow w}.T_{w \leftarrow l_i}.\tau.T_{l_i \leftarrow w}.T_{\overrightarrow{p_i l_i}}.v$$

where T_{proj} is the cone-beam projection and $T_{\text{det} \leftarrow w}$ the transformation matrix from the C-arm world to the detector (X-ray image plane). Both transformations are known because of the C-arm geometry (given in the DICOM file). As the projection of the 3D tip p_i has to match the 2D catheter tip c_1 (i.e. $F_{\text{proj}}(p_i, \tau) = c_1$), the search of the transformation τ translates along the line from c_1 to the origin of the X-ray projection and rotates around it (Fig. 4). l_i is the intersection point of the projection line of c_1 and the plane, containing p_i, parallel to the patient table. It is a coherent starting point for the registration search because we are expecting mostly only a breathing motion of p_i in the cranial-caudal direction. τ is thus computed in the coordinate system around l_i. $T_{\overrightarrow{p_i l_i}}$ is the in-plane translation from p_i to l_i in the world coordinate system.

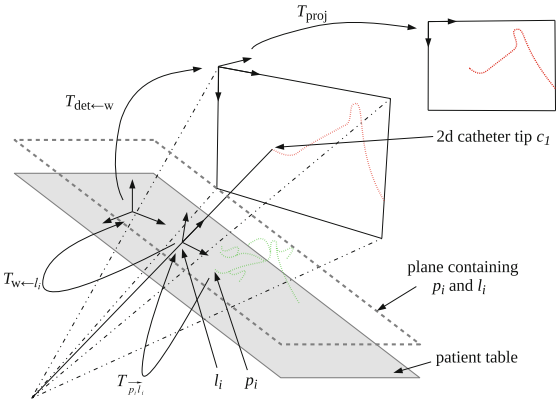


Fig. 4. Transformations in the projection function F_{proj}.

Viterbi. For each image in the sequence, using the Viterbi algorithm [9], the Viterbi path (best tracking of the tip) is computed from the initial state position at the first image to the current image. The result gives the highest probability of the 3D tip position p_i and also (as a 3D/2D registration has been performed during the observations computations) the transformation τ to align the 3D vessel tree inside the 2D fluoroscopy.

3 Experiments and Results

We perform two experiments. First, we evaluate the accuracy of the catheter tip tracking using clinical image data where the catheter and the breathing motion are simulated. In these experiments, the availability of the ground truth from the simulation permits quantitative evaluation. Next, we demonstrate the approach for tip tracking on two real image sequences.

3.1 Implementation

We retrospectively acquired anonymized data of 19 TACE procedures in three different hospitals using intervention rooms with angiographic C-arm systems (Xper Allura, Philips Healthcare, Best, the Netherlands). In total, we acquired 67 fluoroscopic image sequences. A 3DRA was acquired at the beginning of each intervention when the catheter was in the hepatic artery. The 2D catheter is manually segmented in each fluoroscopic image and the 3D vessel tree in the 3DRA is extracted with a semi-automatic method based on thresholding and skeletonization [12].

The matrix A of state transition probabilities is built with $\sigma_a = 12$ mm which enables a relatively large catheter tip motion per frame. The registration score uses $\sigma_s = 1.5$ mm, which penalizes registrations where the normalized sum of the minimal distance $M(C_t, V_i)$ is larger than 1.5 mm. These values have been chosen after a pilot on one simulated image sequence. For the initialization, $\delta_1(i)$ is equal to 1 in the state where the tip is and equal to 0 for all the other states. We manually initialize the tip position of the first image of each sequence. The discretization of the 3D blood vessel and the 2D catheter are set to 3 mm. Because the method has to be computed in real-time, the number of observations $O_t(j)$ to compute at time t is limited to 50. Thus, to calculate every $\delta_t(j)$, we sort in descending order the states s_j based on the probabilities $\max_i[\delta_{t-1}(i) \cdot a_{ij}]$ and compute the observations $O_t(j)$ for the 50 first states s_j in the sorting. All the other state observations are assumed to return a score close to 0, as those states have very low probabilities. Therefore these observation scores are set directly to 0 without any computation and thus δ_t will be also 0. When we compute the observation scores, $M(C_t, V_i)$ is minimized using the Powell optimizer to find the 4 degrees of freedom rigid transformation matrix τ. The search space is limited to $\pm 2°$ for the three rotations and ± 2 mm for the translation along the projection line. The average computation time for each image is below 60 ms with a 2.0 Ghz Intel Core i7 processor.

3.2 Clinical Data with a Simulated Catheter and Breathing Motion

As we do not have ground truth for the catheter position in 3D in our clinical data, we evaluate our method on simulated data. We use the 3D vessel tree from the 3DRA and also the projection geometry of the fluoroscopic images sequence (saved in the DICOM file). Initially, the catheter tip is positioned at a proximal

Table 1. Percentage of incorrect tip tracking at the end of the sequences with different simulation parameters.

Catheter tip speed	No deformation	Slight deformation	Large deformation
1.5 mm/frame	7.8 %	6.2 %	7.8 %
3 mm/frame	4.7 %	1.6 %	0.0 %
6 mm/frame	3.1 %	3.1 %	1.6 %

location in the 3D vessel tree, and over time the tip is advanced. The catheter shape in 3D is obtained by smoothing the corresponding vessel centerline with a Gaussian kernel. The effect of liver motion due to respiration is simulated with a translation along the y-axis (cranial-caudal direction), where the translation magnitude is defined as:

$$translation_y(i) = \lambda \cdot \sin(\frac{2\pi}{\beta} \cdot i \cdot \Delta_t - \frac{\pi}{2})$$

where $\lambda = 10$ mm is the peak amplitude, $\beta = 4$ s the respiration period, i the number of the current image in the sequence and $\Delta_t = 0.133$ s the time between two images, obtained from the fluoroscopic image frequency.

The simulated catheter tip has three different constant speeds: 1.5, 3 and 6 mm/frame (resp. 11.2, 22.5 and 45 mm/s), which is similar to tip speeds during clinical procedures. The simulation has two catheter deformations: slight (Gaussian smoothing with a random (between each image sequence) $\sigma \in [3, 6]$ mm) and large ($\sigma \in [7, 11]$ mm). We tracked the tip and registered all the 67 simulated sequences starting with the correct initial registration at the first frame.

To evaluate the tip tracking, we compute the distance between the real tip and the registered tip chosen by the HMM at every image frame. Both the distance in the 2D image space and the 3D world space are calculated. The average distances between the real 2D projected tip and the registered 2D projected tip for every simulated sequence with different tip speeds and catheter deformations are presented in Fig. 5. The medians of the distances are in a range of 1.1 and 2.7 mm. These distances increase when the deformation is larger and also when the tip moves faster. The average distances between the real 3D tip and the registered 3D tip for every simulated sequence with different tip speeds and catheter deformations are also shown in Fig. 5. The medians of the tip distances are up to 2.9 mm. We define an incorrect tip tracking when the distance between the real 3D tip and the registered 3D tip, in the last 5 images of a tracked sequence, is superior to 3 mm. Following this definition, Table 1 shows the percentage of sequence in which the tracking is incorrect at the end of the sequence. This percentage is up to 7.8 %.

3.3 Clinical Data

We applied the method on two clinical sequences. As there is no ground truth for the 3D tip position available, we qualitatively evaluated whether the tracking is

consistent. In these sequences, the catheter tip is moved from the hepatic artery to a vessel close to the tumor. The first frame is manually registered and the method

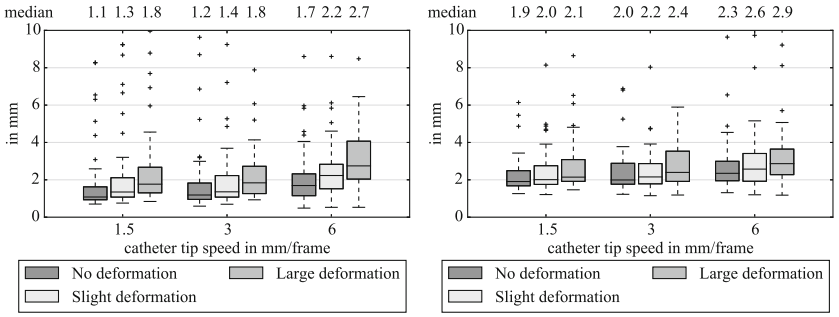


Fig. 5. Average distance for each sequence between the real 2D projected tip and the registered one with different simulation parameters (left). Average distance for each sequence between the real 3D tip and the registered one (right).

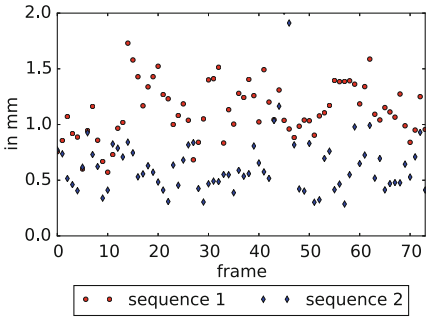


Fig. 6. Average distance (in mm), of the clinical sequences, between points from the 2D catheter centerline and their closest points from the 2D projected vessel centerline V_i, after registration at each frame.

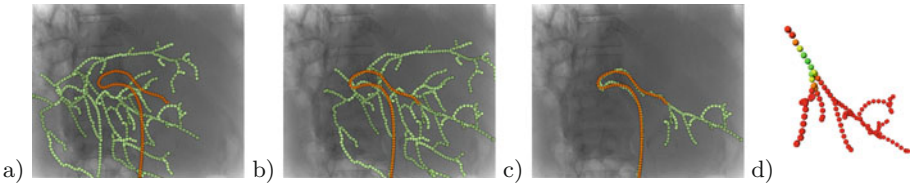


Fig. 7. Tip tracking with 2D catheter (in orange) and 3D vessel tree projection (in green). Before the tracking (a), breathing and table motion prevent the alignment. After the tracking, the roadmap is possible with all the vessel tree visible (b) or only the vessels after the tracked tip (c). Close view of the 3D tracking tip score δ_t inside the vasculature (d) with a colored scale: red (score = 0) to green (score = best score) (Color figure online).

is run on the sequences (74 frames each at 7.5 Hz). Visual checking showed that the registration is approximately correct. Video clips showing the tracking on both sequences are available as supplementary material. In the second sequence, at a bifurcation, the tip goes in the wrong vessel but goes back to the correct one after two frames. We also computed the distance from each catheter point (in 2D) to the closest point of the projected centerline V_i after registration (see Fig. 6). For both sequences, the average point distances between the catheter and the vessel V_i are below 1.9 mm. Figure 7 shows one registered frame of the first sequence.

4 Discussion and Conclusion

We proposed a method for tracking the 3D catheter tip in fluoroscopies using an HMM and registration with the 3D vessel tree extracted from a 3DRA. The method uses the Markov models to estimate the probabilities of the states, which represent the possible catheter tip positions in the 3D vessel tree. The evaluation on simulated data showed a median distance between the real tip and the registered one up to 2.7 mm in the 2D image space and up to 2.9 mm in the 3D space. With the two clinical sequences, we obtain an average distance between the 2D catheter centerline and the projection of the 3D vessel centerline below 1.9 mm.

As the discretization of the catheter centerline and the blood vessel tree is 3 mm, the distances between real tip and registered tip are close to the sampling. The tracking gives better tip accuracy when the motion of the catheter is small and the catheter deformation is slight. The percentage of incorrect tracking at the end of the sequence (up to 7.8 %) is relatively low and is not impacted by deformation. It is higher for the slow catheter motion which implies that the parameter σ_a for the state transition matrix A is probably too large for static and slow catheter motion.

Evaluation of the method on the two clinical cases showed that the catheter and the 3D vessel tree match well. According to our clinical partners, the tracking of the tip position (both in 2D and 3D) is accurate enough to be used for roadmapping and to provide the physician with an overview of where the catheter is and where to go.

Robust automatic 2D catheter segmentation is required after initialization to integrate our method into the interventional workflow. Heibel et al. [4] obtained a mean error of real-time catheter tracking less than 1.2 pixels for abdominal fluoroscopies. Those results should be sufficiently accurate for our registration. In clinical practice, the first initialization of the catheter tip position can be done after the 3DRA acquisition by pointing the catheter tip in the 3DRA. Due to clear visibility of the catheter, this task could be easily automated.

The lack of a ground truth position of the catheter in 3D hampers thorough evaluation of approaches such as ours, which is why we employed extensive experimentation using simulated catheter positions obtained from clinical patient data. Additionally, qualitative evaluation was performed using clinical data only, demonstrating that the method is able to consistently track the position in 3D.

Whereas we have now fixed the parameter σ_a for the state transition matrix A, and the σ_s for converting the registration distance to an observation score, based on a pilot experiment, these values could be tuned to a specific patient or anatomy (e.g. at bifurcations). In the future we therefore intend to investigate the impact of these parameters and how they should be set optimally for each case.

To conclude, we have presented a model to track a catheter tip thanks to 2D fluoroscopies and 3DRA. We evaluated the feasibility of our approach with simulated data demonstrating a tip accuracy below 2.7 mm in 2D image space and 2.9 mm in 3D space. The method was also successfully applied in two clinical cases.

Acknowledgement. This research is funded by Philips Healthcare, Best, The Netherlands.

References

1. Ambrosini, P., Ruijters, D., Niessen, W.J., Moelker, A., van Walsum, T.: Continuous roadmapping in liver tace procedures using 2d–3d catheter-based registration. Int. J. Comput. Assist. Radiol. Surg. 1–14 (2015)
2. Atasoy, S., Groher, M., Zikic, D., Glocker, B., Waggershauser, T., Pfister, M., Navab, N.: Real-time respiratory motion tracking: roadmap correction for hepatic artery catheterizations. In: Medical imaging, pp. 691815–691815. International Society for Optics and Photonics (2008)
3. Groher, M., Zikic, D., Navab, N.: Deformable 2d–3d registration of vascular structures in a one view scenario. IEEE Trans. Med. Imaging **28**(6), 847–860 (2009)
4. Heibel, H., Glocker, B., Groher, M., Pfister, M., Navab, N.: Interventional tool tracking using discrete optimization. IEEE Trans. Med. Imaging **32**(3), 544–555 (2013)
5. Liao, R., Zhang, L., Sun, Y., Miao, S., Chefd'Hotel, C.: A review of recent advances in registration techniques applied to minimally invasive therapy. IEEE Trans. Multimedia **15**(5), 983–1000 (2013)
6. Ma, Y., King, A.P., Gogin, N., Gijsbers, G., Rinaldi, C., Gill, J., Razavi, R., Rhode, K.S.: Clinical evaluation of respiratory motion compensation for anatomical roadmap guided cardiac electrophysiology procedures. IEEE Trans. Biomed. Eng. **59**(1), 122–131 (2012)
7. Markelj, P., Tomaževič, D., Likar, B., Pernuš, F.: A review of 3d/2d registration methods for image-guided interventions. Med. Image Anal. **16**(3), 642–661 (2012)
8. Mitrović, U., Spiclin, Z., Likar, B., Pernuš, F.: 3d–2d registration of cerebral angiograms: a method and evaluation on clinical images. IEEE Trans. Med. Imaging **32**(8), 1550–1563 (2013)
9. Rabiner, L.: A tutorial on hidden markov models and selected applications in speech recognition. Proc. IEEE **77**(2), 257–286 (1989)
10. Rivest-Henault, D., Sundar, H., Cheriet, M.: Nonrigid 2d/3d registration of coronary artery models with live fluoroscopy for guidance of cardiac interventions. IEEE Trans. Med. Imaging **31**(8), 1557–1572 (2012)
11. Ruijters, D., Homan, R., Mielekamp, P., Van de Haar, P., Babic, D.: Validation of 3d multimodality roadmapping in interventional neuroradiology. Phys. Med. Biology **56**(16), 5335 (2011)

12. Selle, D., Preim, B., Schenk, A., Peitgen, H.O.: Analysis of vasculature for liver surgical planning. IEEE Trans. Med. Imaging **21**(11), 1344–1357 (2002)
13. Van Walsum, T., Baert, S.A., Niessen, W.J.: Guide wire reconstruction and visualization in 3dra using monoplane fluoroscopic imaging. IEEE Trans. Med. Imaging **24**(5), 612–623 (2005)

Human-P*n*P: Ergonomic AR Interaction Paradigm for Manual Placement of Rigid Bodies

Fabrizio Cutolo[1(✉)], Giovanni Badiali[3], and Vincenzo Ferrari[1,2]

[1] Department of Translational Research and New Technologies in Medicine and Surgery, EndoCAS Center, University of Pisa, Pisa, Italy
fabrizio.cutolo@endocas.org
[2] Information Engineering Department, University of Pisa, Pisa, Italy
[3] School in Surgical Sciences, University of Bologna, Bologna, Italy

Abstract. The human perception of the three-dimensional world is influenced by the mutual integration of physiological and psychological depth cues, whose complexity is still an unresolved issue per se. Even more so if we wish to mimic the perceptive efficiency of the human visual system within augmented reality (AR) based surgical navigation systems. In this work we present a novel and ergonomic AR interaction paradigm that aids the manual placement of a non-tracked rigid body in space by manually minimizing the reprojection residuals between a set of corresponding virtual and real feature points. Our paradigm draws its inspiration from the general problem of estimating camera pose from a set of n-correspondences, i.e. perspective-*n*-point problem. In a recent work, positive results were achieved in terms of geometric error by applying the proposed strategy on the validation of a wearable AR system to aid manual maxillary repositioning.

Keywords: Augmented reality and visualization · Computer assisted intervention · Interventional imaging

1 Introduction

In the context of image-guided surgery (IGS), augmented reality (AR) technology represents a promising integration between navigational surgery and virtual planning.

In 2012 Kersten-Oertel et al. [1] proposed a taxonomy of mixed reality visualization systems in IGS and defined the three major components based on which they then presented a systematic overview of the trends and solutions adopted in the field [2]. The acronym for the taxonomy (DVV) derives from its three key components: Data type, Visualization Processing and View. According to the taxonomy, for classifying and assessing the efficacy of a new AR system for IGS, we must focus our attention on the particular surgical scenario in which the visualization system aim to be integrated. The surgical scenario affects each of the three DVV factors, namely the type of data that should be displayed at a specific surgical step, the visualization processing technique implemented to provide the best pictorial representation of the augmented scene

C.A. Linte et al. (Eds.): AE-CAI 2015, LNCS 9365, pp. 50–60, 2015.
DOI: 10.1007/978-3-319-24601-7_6

and how and where the output of the visualization processing should be presented to the end-user.

Several visualization processing techniques have been adopted to allow a more immersive viewing experience for the surgeon and a more precise definition of the spatial relationships between real scene and visually processed data along the three dimensions. The human visual system exploits several physiological and psychological cues to deal with the ill-posed inverse problem of understanding a three-dimensional scene from one retinal image. However, monocular and binocular cues are not always sufficient to infer the spatial relationships between objects in the three-dimensional scene. Therefore, a full comprehension of the mechanisms underpinning depth perception is not a completely resolved issue per se in a real scene and it results even more complex within an augmented scene [3]. In this regard, among the suggested visualization processing techniques, researchers have tried to improve the perceptive efficiency by modeling and contextually rendering the virtual content in a photo-realistic manner, and/or by using pixel-wise transparency maps and "virtual windows" [4] to recreate occlusions and motion parallax cues. Some of the proposed techniques for enhancing depth perception comprise high-fidelity texturing [5] or colour coding methods, whereas others consist in lighting and shading cues and/or on the adoption of an interactive "virtual mirror" [6, 7]. Alternatively, depth perception can be improved by relying on standard stereopsis and two-view displays or on more complex full parallax multi-view displays. In any case, to the best of our knowledge, hitherto there are no visualization processing techniques that provide the user with useful information able to improve the postoperative outcome for those specific surgical tasks that involve the accurate manual placement of rigid anatomies in space.

Many surgical procedures in the field of orthopedic surgery or maxillofacial surgery, involve the task of reducing displacements or correcting abnormalities between rigid anatomical structures, i.e. bones, on the basis of a pre-operative planning. The direct tracking of all the rigid anatomies involved in the procedure would yield a measure of the six-degrees-of-freedom displacements between them and it would aid the correct performance of the surgical task, yet it is not always feasible for technical and logistic reasons. In case of single object tracking, the pointer of a standard surgical navigator can be employed by the surgeon to compare the final positions of clearly detectable reference points, over the repositioned anatomy, with those of their counterparts from the surgical planning. Nevertheless, this approach does not allow the assessment of all of the six-degrees-of-freedom at the same time.

AR seems the optimal solution to aid this kind of surgical tasks. Yet, the traditional AR interaction technique featuring the superimposition of a semi-transparent virtual replica of the rigid anatomy in a position and orientation (pose) defined during planning, is not very effective in aiding the surgeon in the correct performance of those procedure. In this regard, it is more beneficial and intuitive for the surgeon to deal with task-oriented visualization techniques, more than with complex reproductions of the virtual anatomies through photorealistic rendering, transparencies and/or virtual windows.

The goal of this work is to present a novel and ergonomic AR interaction paradigm based on a simple visualization processing technique that aims at aiding the accurate manual placement of a non-tracked rigid object in space. Our strategy relies on the tracking of a single object in the scene (e.g. the patient's head), namely on the real-time

estimation of the geometric relation between a scene reference system (SRS) and the camera reference system (CRS), e.g. performed by means of a video based registration approach. In this scenario, the AR guide aids the surgeon in placing other non-tracked rigid bodies (e.g. bones fragments) at a planned pose relative to the CRS. Our paradigm draws its inspiration from the general problem of estimating camera pose from a set of n-correspondences, i.e. perspective-n-point problem. The key idea is that manually minimizing the distance, in the image plane, between a set of corresponding real and virtual feature points is sufficient to aid the accurate placement of a non-tracked rigid body in space.

2 Methods

Perspective-*n*-Point Problem. The task of estimating the pose of a camera with respect to a scene object given its intrinsic parameters and a set of n world-to-image point correspondences is known as the Perspective-*n*-Point (P*n*P) problem in computer vision and exterior orientation or space resection problem in photogrammetry.

This inverse problem concerns many fields of applications (structure from motion, robotics, augmented reality, etc.) and it was first formally introduced in the computer vision community by Fishler and Bolles in 1981 [8], albeit already used in the photogrammetry community before then. According to Fishler and Bolles the P*n*P problem can be defined as follows (distance-based definition):

> *Given the relative spatial locations of n control points* $\boldsymbol{P_i}, \boldsymbol{i} = 1, \ldots n$*, and given the angle to every pair of these points from an additional point called the center of perspective* $\boldsymbol{C}$*, find the lengths* $D_i = |\boldsymbol{CP_i}|$ *of the line segments joining* $\boldsymbol{C}$ *to each of the control points.*

The constraint equations are:

$$D_i^2 + D_j^2 - 2D_iD_j \cos\theta_{ij} = d_{ij}^2, i \neq j \tag{1}$$

Where $D_i = |\boldsymbol{CP_i}|$, $D_j = \left|\boldsymbol{CP_j}\right|$ are the unknown variables, $\theta_{ij} = \widehat{\boldsymbol{P_iCP_j}}$ and $d_{ij} = \left|\boldsymbol{P_iP_j}\right|$ are the known entries (Fig. 1). In computer vision θ_{ij} are determined finding the correspondences between world-to-image points and knowing the intrinsic camera parameters, while d_{ij} are established by the control points.

Following this definition, once each distance D_i is computed, the position of the points $\boldsymbol{P_i}$ can be expressed in the CRS. Therefore, being the position of each $\boldsymbol{P_i}$ in the SRS known, the problem of estimating camera pose with respect to the SRS is reduced to a standard absolute orientation problem whose solution can be found in closed-form fashion through quaternions [9] or singular value decomposition (SVD) [10].

The same problem is also known under the transformation-based definition [11] which can be formalized as:

$$\boldsymbol{\lambda_i \widehat{p}_i} = [\mathrm{K}|0] \begin{bmatrix} \mathrm{R} & \mathrm{T} \\ 0 & 1 \end{bmatrix} \widehat{\boldsymbol{P}}_i, i = 1, \ldots n \tag{2}$$

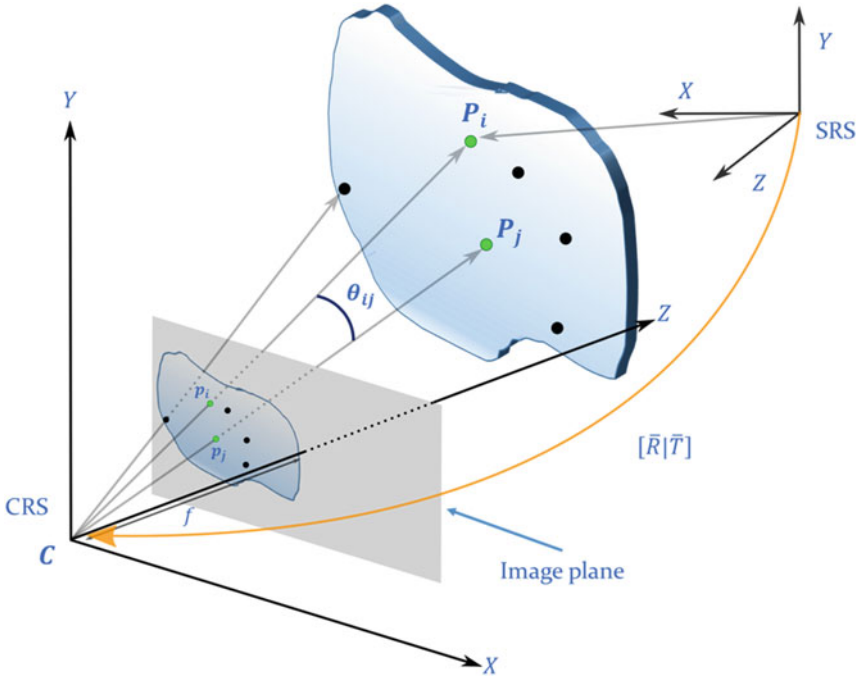


Fig. 1. Geometry of the P*n*P problem.

Where the scene and image points $\hat{\boldsymbol{P}}_{\boldsymbol{i}}$ and $\hat{\boldsymbol{p}}_{\boldsymbol{i}}$ are represented in homogeneous coordinates and the equation is up to a scale factor $\boldsymbol{\lambda_i}$. Hence, according to this definition, the P*n*P problem aims at determining the pose (in terms of a rotation matrix R and a translation vector T) given a set of *n* world-to-image correspondences and known the intrinsic camera parameters encapsulated by the matrix K.

The P*n*P problem has been extensively studied by several groups, which have proposed different iterative, closed-form for solving it.

Closed-form methods [12–18], directly provide an estimation of the camera pose but they are usually less accurate and more susceptible to noise than iterative methods. Iterative non-linear optimization methods solve the P*n*P problem by iteratively minimizing a cost function generally related to the geometric (reprojection residuals) or algebraic error but they need a good initial guess and yield only one solution at a time [19–21]. A useful overview of the state-of-the-art methods can be found in [17] and in [22].

In terms of geometric reprojection residual, the non-linear cost function can be formulated as the sum of the squared measurement errors (d_i):

$$\begin{aligned} \bar{\mathrm{R}}|\bar{\mathrm{T}} &= \arg\min \sum_{i=1}^{n} d(\boldsymbol{p_i}, \hat{\boldsymbol{p}}_i)^2 \\ &= \arg\min \sum_{i=1}^{n} \left\| \boldsymbol{p_i} - \hat{\boldsymbol{p}}_i(\mathrm{K}, \hat{\mathrm{R}}, \hat{\mathrm{T}}, \boldsymbol{P_i}) \right\|^2 \end{aligned} \tag{3}$$

Where $\boldsymbol{p_i}$ are the measured image points, and $\hat{\boldsymbol{p}}_{\boldsymbol{i}}$ are the calculated projections of the corresponding control points as a function of $\mathrm{K}, \hat{\mathrm{R}}, \hat{\mathrm{T}}$.

The other important research direction on the P*n*P problem is the study of the multi-solution phenomenon of the P*n*P problem [23], principally when $n = 3$ (P3P) [24, 25], being three the smallest subset of control points that yields a finite number of solutions. P3P problem yields at most four solutions which can be disambiguated using a fourth point, and it is the most studied case since it can be used as first step to reduce the complexity of the computation of a P*n*P problem, e.g. in a RANSAC scheme by removing the outliers.

AR Video-Based Camera Registration. Regardless of the method adopted for solving the P*n*P problem, an immediate application of the P*n*P problem is to locate the pose of a calibrated camera with respect to an object, given the 3D position of a set of n control points rigidly constrained to the object and the 2D position of their correspondent projections onto the image plane.

For a correct registration of computer-generated elements to the real scene in AR-based surgical navigation systems, the image formation process of the virtual camera must perfectly mimic the real camera one. In mostly all the AR applications the estimation of the intrinsic camera parameters is the result of an off-line calibration process whereas the extrinsic camera parameters are determined online, e.g. solving a P*n*P problem in real-time. This video-based camera registration method suggested us the implementation of an ergonomic AR interaction paradigm for positioning and orienting a non-tracked rigid object in space.

Human-P*n*P. As written in the introduction, many surgical procedures in the field of orthopedic surgery or maxillofacial surgery, involve the task of manually placing rigid anatomies on the basis of preoperative planning. In that case, let us assume that we can rely on a robust and accurate registration of the surgical planning onto the real scene, by means of the tracking of at least one rigid body (e.g. the head). The six-degrees-of-freedom pose of an additional and non-tracked rigid anatomy in relation to the SRS, can be retrieved by physically placing it as to minimize the geometric distance, on the image plane, between a set of real and virtual feature points. For brevity, from now on, we shall refer to these structures as "tracked anatomy" for the former and "non-tracked anatomy" for the latter, while the proposed method will be referred to as the human-perspective-*n*-point problem (hP*n*P).

From a theoretical standpoint, our method draws its inspiration and physically mimics the paradigm on which the P*n*P problem is formulated. As mentioned in the previous section, the main goal of the P*n*P problem is to infer useful information on the real 3D scene, based on 2D observations of it. In an AR application, this spatial information is used to geometrically register the virtual elements onto the real scene. Thus, as a general rule and regardless of the method adopted for solving the P*n*P problem, a robust and accurate registration should minimize in the image plane the geometric reprojection residuals between measured and estimated projections (see Eq. 3). Similarly, the goal of our hP*n*P interaction paradigm is to achieve the desired placement of a non-tracked anatomy by manually minimizing the reprojection residuals

between correct/planned projections $\bar{\boldsymbol{p}}_{\boldsymbol{i}}$ of virtual landmarks, and observed projections $\hat{\bar{\boldsymbol{p}}}_{\boldsymbol{i}}$ of real landmarks.

The correct/planned projections $\bar{\boldsymbol{p}}_{\boldsymbol{i}}$ are rendered on the image plane according to the real-time estimation of the camera pose $[\bar{\mathrm{R}}, \bar{\mathrm{T}}]$ relative to the tracked anatomy reference system (SRS) and assuming the intrinsic camera parameters, encapsulated by matrix K, are determined offline, e.g. through the Zhang's method [26]. The position of each virtual landmark $\boldsymbol{P}_{\boldsymbol{i}}$ in the SRS is established during surgical planning.

The real projections $\hat{\bar{\boldsymbol{p}}}_{\boldsymbol{i}}$ are associated with the pose, encapsulated by $[\hat{\bar{\mathrm{R}}}, \hat{\bar{\mathrm{T}}}]$, between viewing camera and non-tracked anatomy reference frame (ARS): this resulting pose varies according to the manual placement of the rigid body relative to the camera:

$$\begin{aligned} \tilde{\mathrm{R}}|\tilde{\mathrm{T}} &= \arg\min \sum_{i=1}^{n} d\left(\bar{\boldsymbol{p}}_{\boldsymbol{i}}, \hat{\bar{\boldsymbol{p}}}_{\boldsymbol{i}}\right)^2 \\ &= \arg\min \sum_{i=1}^{n} \left\| \bar{\boldsymbol{p}}_{\boldsymbol{i}}(\mathrm{K}, \bar{\mathrm{R}}, \bar{\mathrm{T}}, \boldsymbol{P}_{\boldsymbol{i}}) - \hat{\bar{\boldsymbol{p}}}_{\boldsymbol{i}}(\mathrm{K}, \hat{\bar{\mathrm{R}}}, \hat{\bar{\mathrm{T}}}, \boldsymbol{P}_{\boldsymbol{i}}) \right\|^2 \end{aligned} \tag{4}$$

In this way, we wish to obtain $\left[\tilde{\mathrm{R}}|\tilde{\mathrm{T}}\right] \approx [\bar{\mathrm{R}}|\bar{\mathrm{T}}]$ (see Fig. 2), namely we seek to positioning and orienting the ARS as coincident with the planned and registered SRS (non-tracked anatomy reference frame ≈ planning reference frame).

To implement this strategy, we add simple virtual elements (e.g. virtual asterisks, crosses, etc.) to the virtual scene during the surgical planning: one element for each of the clearly detectable physical landmarks on the rigid body. The landmarks may consist of a series of distinguishable feature points over the surface of the anatomy or rigidly

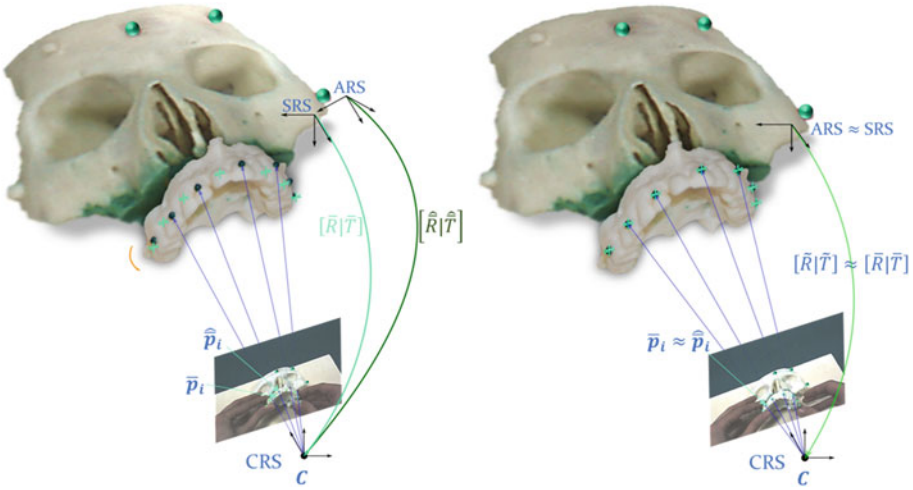


Fig. 2. Geometry of the hP*n*P: minimizing the reprojection residual between registered projections $\bar{p}_i$ and real projections $\hat{\bar{p}}$ is sufficient to aid the accurate placement of a rigid body (the maxilla in the image) in space.

constrained to it. Under such AR guidance, the user moves the non-tracked rigid body up to obtain a perfect overlapping between real and virtual landmarks, hence manually minimizing the reprojection residuals on the image plane: $\bar{\boldsymbol{p}}_{\boldsymbol{i}} \approx \hat{\bar{\boldsymbol{p}}}_{\boldsymbol{i}} \forall \boldsymbol{i}$ (Fig. 3). The theoretical assumptions underpinning the P*n*P problem ensure that if $\bar{\boldsymbol{p}}_{\boldsymbol{i}} \approx \hat{\bar{\boldsymbol{p}}}_{\boldsymbol{i}} \forall \boldsymbol{i}$, the non-tracked anatomy is placed in the correct pose as planned in SRS.

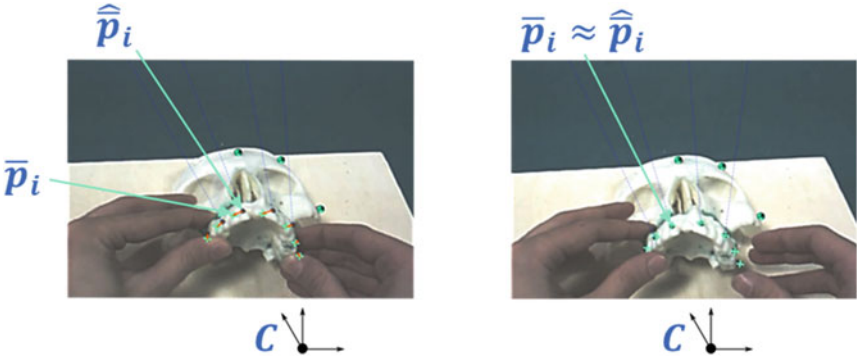


Fig. 3. Detail of the image plane with the minimization of the reprojection residuals. Here the virtual information consists of a cyan-colored asterisk for each physical landmark clearly detectable over the maxilla, e.g. coloured landmarks fixed on the brackets of the orthodontic appliance (Color figure online).

3 Results

In a recent work [27], the described strategy was applied in the validation of a wearable AR system to aid maxillary repositioning. AR system consisted of a stereoscopic video see-through head mounted display equipped with two external USB cameras placed in a quasi-orthoscopic position [28, 29]. The video see-through paradigm of the system is implemented as follows (Fig. 4): real-world views are grabbed by a pair of calibrated external cameras; the captured frames, after compensation of the radial distortion, are screened as backgrounds of the virtual scene onto the corresponding display; the virtual elements, defined during planning, are added to the real scene and observed by a pair of virtual cameras whose processes of image formation mimic those of the real cameras in terms of intrinsic and extrinsic camera parameters. Zhang's method is used to calibrate the two cameras. The estimation of the extrinsic parameters, allowing the real-time registration of the virtual elements to real scene, is achieved through a marker-based video-registration method [29].

In the study, manual repositioning of the upper maxilla following LeFort 1 osteotomy was chosen as test procedure. The test was conducted on a CT-scanned/3D-printed replica of a cadaveric human skull. The planned pose of the maxilla, as defined during preoperative planning, acts as a guide for the surgeon during the intervention performed in-vitro. The traditional AR interaction technique, featuring the superimposition of a semi-transparent virtual replica of the maxilla, as dictated by the surgical planning, did

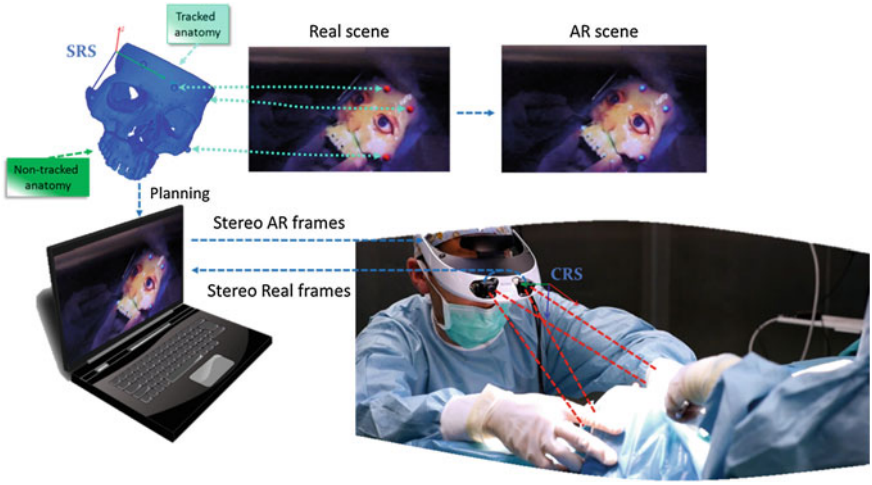


Fig. 4. Video see-through paradigm of the stereoscopic head mounted display used to aid maxillary repositioning.

not prove to be very effective in aiding the surgeon in manually repositioning the upper maxilla. This was mostly due to the surgeon's limited perception of the relative distances of objects within the AR scene owing to the presence of unnatural occlusions between the real and the virtual maxilla. Conversely, a more ergonomic form of visualization consisted in the use of an interaction paradigm which actualizes the above described hP*n*P approach: physical landmarks onto the maxilla and corresponding to coloured landmarks fixed on the brackets of the orthodontic appliance usually applied prior to this kind of interventions, were designated as reference markers for the AR view modality. The repositioning of the maxilla is assisted by visually aligning small virtual asterisks, drawn in positions defined during planning (relative to the SRS), with the corresponding real landmarks.

The upper surface of the maxilla (corresponding to the post-osteotomy surface) was covered with highly malleable plasticine so to be fixed to the upper skull once the surgeon performed the repositioning. The surgical accuracy was validated with the aid of an optical navigation system that recorded the coordinates of three reference points on the non-tracked maxilla after repositioning. Six surgeons and three unskilled engineers were involved in the testing, each of whom was asked to manually reposition the maxilla as dictated by three surgical plannings of variable complexity. Results in terms of linear distances between the real positions of the reference holes and the expected positions (defined during planning) were very promising: mean error was 1.70 ± 0.51 mm. The axial errors were 0.89 ± 0.54 mm on the sagittal axis, 0.60 ± 0.20 mm on the frontal axis, and 1.06 ± 0.40 mm on the cranio-caudal axis. Such results were obtained without the tracking of the maxilla but just relying on the ergonomics of the chosen AR interaction paradigm: the overlapping on the image plane between virtual feature points and real landmarks, visible over the non-tracked anatomy, proved to be sufficient to aid the accurate repositioning of the maxilla.

4 Discussion

It is important to note that the chosen AR interaction paradigm was not bound to the particular video-based tracking modality exploited in the cited study, neither to the use of a specific wearable stereoscopic system. Howbeit, the user can enhance the accuracy in object placement by checking consistency of real and virtual landmarks from different viewpoints. In this regard, the ergonomics of the proposed method may benefit from the adoption of a wearable AR system. Moreover, the choice of such instance of visualization data was, in that work, empirically inspired by the authors' endeavor of defining a modality that were ergonomic for the surgeon and that provided the smallest perceived parallax error: no further discussion was held on the theoretical hypotheses behind such interaction paradigm which are here discussed for the first time.

5 Conclusion

In this work, we proposed a novel and ergonomic AR interaction paradigm that aims at obtaining the accurate placement of a rigid body in space without the need for multiple objects tracking and/or complex visual representations of the virtual guide. From a theoretical standpoint, our method draws its inspiration and physically mimics the paradigm on which the P*n*P problem in computer vision is formulated. This approach, represented by the acronym hP*n*P, could be of help in those tasks, also not specifically surgical, where the AR guide aims at aiding the placement of a rigid body in space. The key-principle behind this interaction paradigm can be exploited in many different AR-based navigation systems: it can be integrated with different end-products of the visualization process in terms of display technology and perception location and/or it could be realized in conjunction with various tracking modalities.

To increase robustness and applicability of the proposed AR interaction paradigm in a real clinical scenario, redundancy in choosing the set of landmarks must be granted. Further, the presence of line-of-sight occlusions caused by soft-tissues, surgeon's hands or surgical instrumentation may be restricted by conveniently selecting the position of the landmarks in relation to the surgical field.

Acknowledgments. This work was funded by the Italian Ministry of Health grant SThARS (Surgical training in identification and isolation of deformable tubular structures with hybrid Augmented Reality Simulation, 6/11/2014–5/11/2017). Grant "Ricerca finalizzata e Giovani Ricercatori 2011–2012" Young Researchers – Italian Ministry of Health.

References

1. Kersten-Oertel, M., Jannin, P., Collins, D.L.: DVV: towards a taxonomy for mixed reality visualization in image guided surgery. In: Liao, H., Edwards, P.J., Pan, X., Fan, Y., Yang, G.-Z. (eds.) MIAR 2010. LNCS, vol. 6326, pp. 334–343. Springer, Heidelberg (2010)
2. Kersten-Oertel, M., Jannin, P., Collins, D.L.: The state of the art of visualization in mixed reality image guided surgery. Comput. Med. Imaging Graph. **37**, 98–112 (2013)

3. Sielhorst, T., Bichlmeier, C., Heining, S.M., Navab, N.: Depth perception - a major issue in medical AR: evaluation study by twenty surgeons. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4190, pp. 364–372. Springer, Heidelberg (2006)
4. Bichlmeier, C., Wimme, F., Heining, S.M., Navab, N.: Contextual anatomic mimesis hybrid in-situ visualization method for improving multi-sensory depth perception in medical augmented reality. In: 6th IEEE and ACM International Symposium on Mixed and Augmented Reality, ISMAR 2007, pp. 129–138 (2007)
5. Haouchine, N., Dequidt, J., Berger, M.O., Cotin, S.: Single view augmentation of 3D elastic objects. In: 2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), pp. 229–236 (2014)
6. Bichlmeier, C., Heining, S.M., Feuerstein, M., Navab, N.: The virtual mirror: a new interaction paradigm for augmented reality environments. IEEE Trans. Med. Imaging **28**, 1498–1510 (2009)
7. Bichlmeier, C., Euler, E., Blum, T., Navab, N.: Evaluation of the virtual mirror as a navigational aid for augmented reality driven minimally invasive procedures. In: 2010 9th IEEE International Symposium on Mixed and Augmented Reality (ISMAR), pp. 91–97 (2010)
8. Fischler, M.A., Bolles, R.C.: Random sample consensus - a paradigm for model-fitting with applications to image-analysis and automated cartography. Commun. ACM **24**, 381–395 (1981)
9. Horn, B.K.P.: Closed-form solution of absolute orientation using unit quaternions. J. Opt. Soc. Am. A-Opt. Image Sci. Vis. **4**, 629–642 (1987)
10. Arun, K.S., Huang, T.S., Blostein, S.D.: Least-squares fitting of 2 3-D point sets. IEEE Trans. Pattern Anal. Mach. Intell. **9**, 699–700 (1987)
11. Wu, Y.H., Hu, Z.Y.: PnP problem revisited. J. Math. Imaging Vis. **24**, 131–141 (2006)
12. Haralick, R.M., Lee, C.N., Ottenberg, K., Nolle, M.: Review and analysis of solutions of the 3-point perspective pose estimation problem. Int. J. Comput. Vis. **13**, 331–356 (1994)
13. Quan, L., Lan, Z.D.: Linear N-point camera pose determination. IEEE Trans. Pattern Anal. Mach. Intell. **21**, 774–780 (1999)
14. Fiore, P.D.: Efficient linear solution of exterior orientation. IEEE Trans. Pattern Anal. Mach. Intell. **23**, 140–148 (2001)
15. Gao, X.S., Hou, X.R., Tang, J.L., Cheng, H.F.: Complete solution classification for the Perspective-Three-Point problem. IEEE Trans. Pattern Anal. Mach. Intell. **25**, 930–943 (2003)
16. Ansar, A., Daniilidis, K.: Linear pose estimation from points or lines. IEEE Trans. Pattern Anal. Mach. Intell. **25**, 578–589 (2003)
17. Lepetit, V., Moreno-Noguer, F., Fua, P.: EPnP: an accurate O(n) solution to the PnP problem. Int. J. Comput. Vis. **81**, 155–166 (2009)
18. Hesch, J.A., Roumeliotis, S.I.: A direct least-squares (DLS) method for PnP. In: 2011 IEEE International Conference on Computer Vision (ICCV), pp. 383–390 (2011)
19. Haralick, R.M., Joo, H., Lee, C.N., Zhuang, X.H., Vaidya, V.G., Kim, M.B.: Pose estimation from corresponding point data. IEEE Trans. Syst. Man Cybern. **19**, 1426–1446 (1989)
20. Lowe, D.G.: Fitting parameterized 3-Dimensional models to images. IEEE Trans. Pattern Anal. Mach. Intell. **13**, 441–450 (1991)
21. Lu, C.P., Hager, G.D., Mjolsness, E.: Fast and globally convergent pose estimation from video images. IEEE Trans. Pattern Anal. Mach. Intell. **22**, 610–622 (2000)
22. Garro, V., Crosilla, F., Fusiello, A.: Solving the PnP problem with anisotropic orthogonal procrustes analysis. In: Second Joint 3dim/3dpvt Conference: 3d Imaging, Modeling, Processing, Visualization & Transmission (3dimpvt 2012), pp. 262–269 (2012)

23. Hu, Z.Y., Wu, F.C.: A note on the number of solutions of the noncoplanar P4P problem. IEEE Trans. Pattern Anal. Mach. Intell. **24**, 550–555 (2002)
24. Zhang, C.X., Hu, Z.Y.: Why is the danger cylinder dangerous in the P3P problem. Zidonghua Xuebao/Acta Automatica Sinica **32**, 504–511 (2006)
25. Wang, T., Wang, Y.C., Yao, C.: Some discussion on the conditions of the unique solution of P3P problem. IEEE ICMA 2006: Proceeding of the 2006 IEEE International Conference on Mechatronics and Automation, vols. 1–3, Proceedings, pp. 205–210 (2006)
26. Zhang, Z.Y.: A flexible new technique for camera calibration. IEEE Trans. Pattern Anal. Mach. Intell. **22**, 1330–1334 (2000)
27. Badiali, G., Ferrari, V., Cutolo, F., Freschi, C., Caramella, D., Bianchi, A., Marchetti, C.: Augmented reality as an aid in maxillofacial surgery: validation of a wearable system allowing maxillary repositioning. J. Craniomaxillofac. Surg. **42**, 1970–1976 (2014)
28. Ferrari, V., Megali, G., Troia, E., Pietrabissa, A., Mosca, F.: A 3-D mixed-reality system for stereoscopic visualization of medical dataset. IEEE Trans. Biomed. Eng. **56**, 2627–2633 (2009)
29. Cutolo, F., Parchi, P.D., Ferrari, V.: Video see through AR head-mounted display for medical procedures. In: 2014 IEEE International Symposium on Mixed and augmented reality (ISMAR), pp. 393–396 (2014)

Real-Time Markerless Respiratory Motion Management Using Thermal Sensor Data

Hagen Kaiser(✉), Pascal Fallavollita, and Nassir Navab

Technische Universität München, Munich, Germany
hagen.kaiser@brainlab.com

Abstract. We present a novel approach to jointly generate multidimensional breathing signals and identify unintentional patient motion based on thermal torso imaging. The system can operate at least 30 % faster than the currently fastest optical surface imaging systems. It provides easily obtainable point-to-point correspondences on the patients surface which makes the current use of computationally heavy non-rigid surface registration algorithms obsolete in our setup, as we can show that 2d tracking is sufficient to solve our problem. In a volunteer study consisting of 5 patient subjects we show that we can use the information to automatically separate unintentional movement, due to pain or coughing, from breathing motion to signal the user that it is necessary to re-register the patient. The method is validated on ground-truth annotated thermal videos and a clinical IR respiratory motion tracking system.

1 Introduction

In external beam radiation therapy it is crucial to keep the patient aligned with respect to pre-operative planning data during the whole treatment fraction. The task can be separated into two categories: setup and motion management. While setup tries to align the patient with the highest possible accuracy, motion management tries to keep the patient aligned over time. Motion management is two-fold if treatment is applied to extra-cranial sites: respiratory motion, which can be used as a surrogate signal for internal tumor movement, and spontaneous movement of the patient. Range Imaging (RI) sensors promise to provide continuous, non-invasive imaging and therefore have been studied for motion management [6,7]. Typical approaches try to separate respiratory motion from patient movement. Early versions of the AlignRT system (Vision RT, London UK) acquired RI surfaces for one respiratory cycle. A respiratory signal was extracted and mapped to the surfaces. Afterwards only surfaces recorded during End of Exhale were registered to each other, which inevitably drops the frame rate to 0.1–0.3 Hz. An extension to this idea was proposed by Wasza et al. [7], where the whole respiratory cycle is used to build a deformation model that later can be used for an efficient non-rigid registration. In a following work [8] a multi-RI sensor fusion framework was developed to increase the patient's body coverage since linear independence of the deformation vectors of non-rigidly registered surfaces potentially allows a better separation of rigid and non-rigid

C.A. Linte et al. (Eds.): AE-CAI 2015, LNCS 9365, pp. 61–68, 2015.
DOI: 10.1007/978-3-319-24601-7_7

movement. This goes in line with [3] who outline general problems of 3D surface registration algorithms, which suffer from a decrease of robustness near or on featureless areas, like thorax or abdomen. All these works try to separate non-rigid from rigid movement due to the assumption that non-rigid movement is equivalent to respiratory motion and rigid-movement to spontaneous movement which then can be compensated by a robotic couch. These assumptions are inaccurate since a patients movement on a couch is always highly non-rigid. Some tumor tracking devices are capable to work with very densely sampled surrogate signals at 60 Hz (Vero System, Brainlab). Such tracking rates currently can only be achieved by marker based systems. To our knowledge the fastest known RI system achieves 20 Hz [4]. This is due to the complex nature of structured light systems, which have to project one or more patterns per frame on the patient's body with a camera filming the patterns in exact synchronization. This technology is somehow limited since to increase frame-rate also an increased amount of light is needed for the pattern projection, which at some point can become dangerous for the patients retina.

Contributions. In this paper, we propose an algorithm which separates non-rigid respiratory motion from non-rigid spontaneous movement while being able to generate multi-dimensional breathing signals to serve as a surrogate for tumor motion compensation. We also make use of thermal imaging which only relies on the thermal energy emanated by the patients body. The lower priced system used in our work runs comfortably at 30 Hz. From this, we exploit the fact that the patients torso consists of many distinctive thermal regions exactly at those locations where typically used state-of-the-art RI sensors give a low signal. These features can be tracked directly in the 2D sensor plane which makes the use of computational heavy non-rigid surface algorithms obsolete in our setup. Although this MICCAI paper tracks surface deformation only implicitly, the assumption also would hold if used in a setup that reconstructs the 3D surface based on motion or stereo using thermal imaging.

2 Methods

Our method relies on thermal image sensors delivering spatially consistent thermal surface information over time which can be approximated in a simplified manner: a scene point $X \in \mathbb{R}^3$ on a potentially moving surface appears in the thermal sensor-plane as

$$\Theta(x) := T(X), \text{with } x = PX \tag{1}$$

where P projects X from scene coordinate system to $x \in \mathbb{R}^2$ in the two-dimensional sensor plane, and $T(X) \in R$ represents the temperature of the measured surface point. In our setup we film a patients naked chest and abdominal region with an acquisition rate of 30 Hz. Given two successive thermal measurements at time t and $t+1$ of a moving surface location, $X_t = (x, y, z)$ could

move to $X_{t+1} = (x + \Delta x, y + \Delta y, z + \Delta z)$, but assuming the body is not rapidly cooling down or heating up at the given location yields

$$T(X_t) \approx T(X_{t+1}) \tag{2}$$

The obvious nice property of 2 is its independence of external lighting conditions or patients skin color. Equations 1 and 2 give way to a reformulation of the well known *brightness constancy assumption* as a *temperature constancy assumption*: Let $\Theta(x, y, t)$ be a filmed sequence of thermal images and let $(x(t), y(t))$ be a trajectory of a point in the thermal sensor plane then:

$$\frac{d}{dt}\Theta(x(t), y(t), t) = 0 \tag{3}$$

In the results section we will see that this assumption holds in a stable room climate, which is traditionally the case in all radiation therapy treatment rooms. Following the assumption of 1 such trajectories directly relate to the deformation of the patients chest surface. Given that a chest surface consists of sufficiently distinctive thermal regions we can estimate a vector field u that matches a point in one image to a shifted point in another such that:

$$\nabla\Theta * u + \frac{\delta}{\delta t}\Theta = 0 \tag{4}$$

To estimate u we employ [1], which implements a pyramidal block matching algorithm which minimizes:

$$\sum_x [\Theta(W(x; u)) - J(x)]^2 \tag{5}$$

where $J(x)$ is an extracted subregion of the sensor measurement at $t = 1$ and $\Theta(x)$ is the thermal image at $t = 2$. $W(x; u)$ warps each pixel from $J(x)$ to sub-pixel location in $\Theta(x)$. To find subregions that can be tracked with high stability we detect corner like features around the borders of equithermal regions [5].

2.1 An External Surrogate for Tumor Tracking

By tracking thermal landmarks on the patient's surface we get an implicit description of the body surface deformation. Current radiation therapy systems often use a single-dimensional surrogate signal (e.g. to gate the treatment beam). To investigate the applicability of our approach we demonstrate how we can build a meaningful single or low-dimensional breathing signal out of the tracked landmarks. After the patient has been setup for treatment we build a respiratory motion model which we can use for dimensionality reduction. The model acquisition process is started by the user marking at least one region of interest (ROI) in the live thermal video stream. By key-press the user can now capture at least one full EI (End of Inhale) to EE (End of Exhale) breathing phase measured by n frames. For each ROI a feature detection is performed to find $c := 1, 2, .., j$

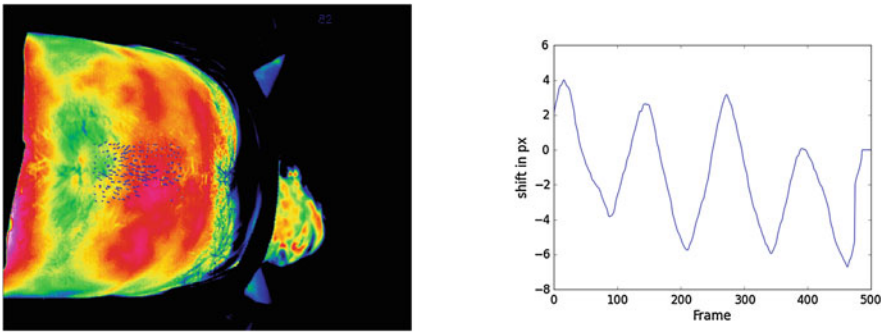


Fig. 1. Respiratory signal generation by analyzing the optical (thermal) flow

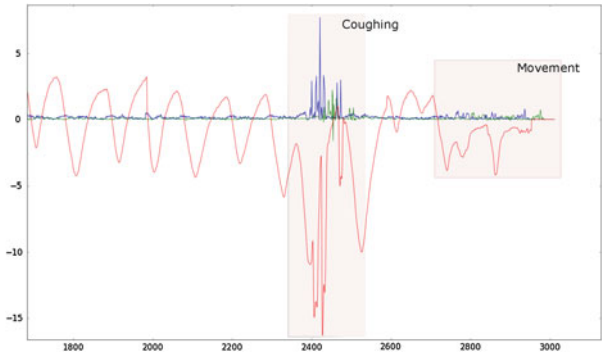


Fig. 2. Red: Respiratory signal b_{ROI}; (x:=FrameNr, y:=Shift in px) Green: Directional Change ϕ_t (x:=FrameNr, y:=angle) Blue: Flow Velocity $\Delta\Gamma_t$ (x:=FrameNr, y:=Px/Frame) (Color figure online)

landmarks suitable for tracking. For a tracked landmark we build a track of measurements $\Gamma_1 c, \Gamma_2 c, ..., \Gamma_n c \in \mathbb{R}^2$ (Fig. 1).

We do a Principal Component Analysis on each track c and extract the largest Eigenvector e_c and project, then each incoming track is updated from Γ_t.

$$b_{\text{ROI}}(t) = \frac{1}{j} \sum e_c \Gamma_t \tag{6}$$

Equation 6 yields a one dimensional breathing signal for a ROI. This is similar to state of the art marker based breathing signal trackers as Vero [2], which average the anterior-posterior (AP) movement of several markers around a region of interest.

2.2 Separate Patient Movement from Respiratory Motion

To separate unintended patient movement from respiratory motion we build a model out of the fist few seconds of respiratory motion tracking. Respiratory

motion can be considered as a *slow* motion. In our preliminary experiments when using the ground-truth Vero System (i.e. marker-based) we observed that each marker typically moves around 2–3 mm in three seconds. Any other form of human motion, whether it be coughing and patient movement on the couch, was considerably faster. Analyzing our videos we could also identify a sudden change in track direction as a typical feature (see Fig. 2). After having recorded one respiratory cycle we transform all measurements $\Gamma_{1c}, \Gamma_{2c}, ..., \Gamma_{nc}$ for each track to polar coordinates $\Pi 1c, \Pi 2c, ..., \Pi nc$ with

$$\Pi := (r, \phi), \text{with } \Gamma(\Pi) := (r \cos \phi, r \sin \phi) \tag{7}$$

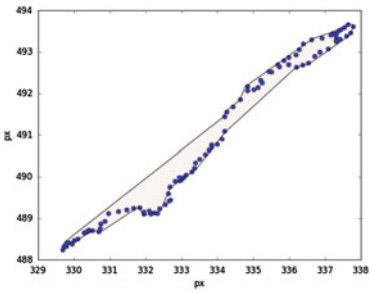


Fig. 3. Approximation of respiratory motion in polar coordinates. The margin can be scaled by tuning parameter f_d

We then build a lookup-table LUT_c in which we save for each rounded angle $\psi(\phi) \approx \phi :\in 0, 1, .., 360$ the maximum radius r discovered through the respiratory cycle (see Fig. 3. Furthermore we calculate the average flow velocity $v_c := \sum_1^n \Delta\Gamma$. We then use a classification function on each track to decide if a current measurement signals *respiratory motion* 0 or *movement* 1:

$$M_c(\Gamma_c t, f_v, f_d) := \begin{cases} 1 \text{ if } ||\Delta\Gamma_c t - v_c|| > f_v \sigma_c \\ 1 \text{ if } f_d \text{ LUT}(\psi(\phi_{tc})) > r_{tc} \\ 0 \qquad \text{otherwise} \end{cases} \tag{8}$$

whereas σ_{tc} is the standard deviation of the track's average flow velocity v_c, ϕ_{tc} and r_{tc} are the angular and radial component of Π_{tc} respectively and f_v and f_d are tuning parameters to increase or decrease tolerance. To separate respiratory motion from spontaneous movement, we classify each track with its respective current measurement and decide by majority vote.

3 Results

Study Protocol. The evaluation was performed using an InfraTec Vario Cam HD 900 Research (30 Hz, Resolution 1024×768). The tuning parameters were

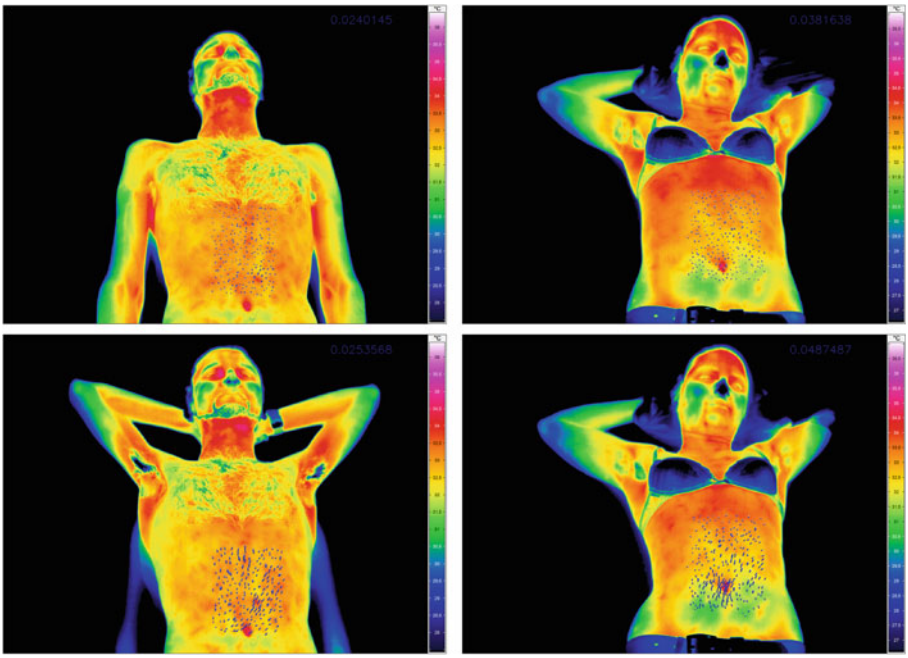


Fig. 4. (Top row) Breathing tracks during respiratory motion; (Bottom row) breathing tracks due to coughing or movement

chosen as $f_v = 2$ and $f_d = 1.5$. While f_v was chosen typically for outlier detection using the standard deviation, f_d was chosen intuitively but generalized well for all tracked patients (Fig. 4). The tracking algorithm worked on thermal subregions of 16×16 pixels. We performed a volunteer study on 4 male and 1 female subject. Each patient had their breathing tracked for 2 min. Figure 5 depicts the different breathing patterns seen on each patient. To investigate the quality of the respiratory signals we compared the signal of a single b_{ROI} with the ground-truth signal of the marker-based motion tracker of a VERO Linear Accelerator (Brainlab, Germany). We restricted the VERO Tracker to use a single marker. We explicitly masked the trajectory area of the marker, that was in the center of our ROI, to not confuse our tracking algorithm. During the experiments, the patients were instructed to randomly cough and subtly move to produce diverse signals. One can see in Fig. 5 that although the signals are not aligned (as expected) they have a remarkable congruity. We repeated recordings with each patient for another two minutes but without any marker. Here we tested the quality of our movement and coughing classifier against the ground-truth manual annotation: The frames were carefully labeled with respect to the start of movement or coughing. We then applied our classification algorithm to each video. As ROI we used a user defined rectangular region around the solarplexus. On average 243 trackable thermal features were found. The tracking algorithm

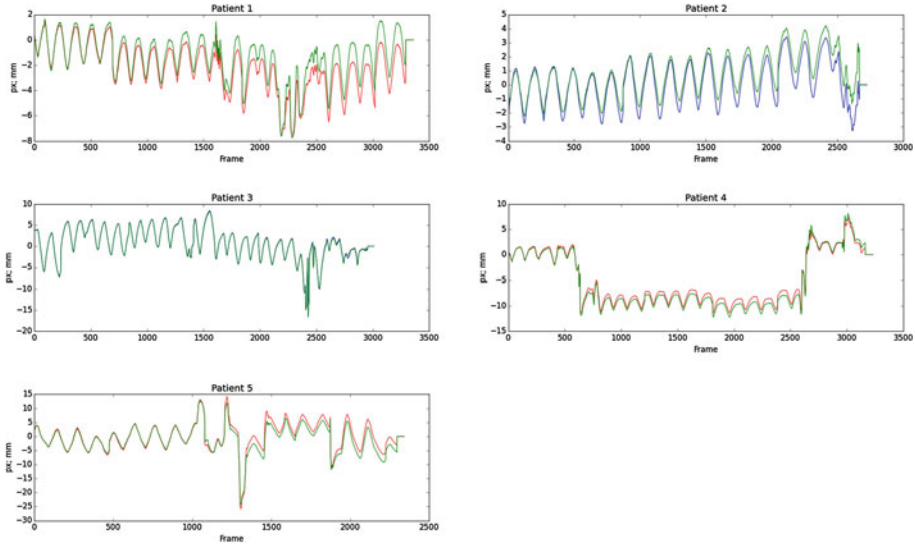


Fig. 5. Respiratory signals. Red: thermal tracking; Green: VERO marker-based tracking (Color figure online)

including classification was implemented without GPU acceleration on an Intel Xeon E5-1650 3.5 Ghz under 28 ms/iteration. To build the model we captured by key-press one whole EE to EI respiration phase in around 3 to 6 s. The classifier was able to predict all movements and coughing in all videos in a range of ± 7 frames which is acceptable considering a user labeled sequence. It is also noteworthy that the experiments were performed in temperature-controlled treatment room with 22°. The camera pictures were thresholded between 25° and 35° which is a cheap way of segmenting the patient's skin from all other surroundings yielding sufficiently many thermal regions. When thresholding was done with a range below 5° the pictures began to suffer from severe noise artifacts.

4 Conclusion

We presented a marker-less real-time framework for respiratory motion analysis and motion and management. As our major contribution we proposed to use thermal cameras and track torso deformation implicitly by analyzing the optical/thermal flow of equithermal regions on the patient's body. We furthermore demonstrated a simple but effective model to jointly analyze the respiratory motion and classify the detected moving patterns as respiratory or spontaneous patient motion, such that a beam-off signal could be triggered automatically during a radiation therapy treatment, which could significantly increase patient safety. Due to the high frame rates of off-the-shelve thermal cameras this framework is eligible to be integrated into tumor tracking setups in radiation therapy, where a marker-less solution can significantly reduce setup time and increase reproducibility between fractions.

References

1. Bouguet, J.-Y.: Pyramidal implementation of the Lucas Kanade feature tracker description of the algorithm. Practice **1**(2), 1–9 (1999)
2. Depuydt, T., Verellen, D., Haas, O., Gevaert, T., Linthout, N., Duchateau, M., Tournel, K., Reynders, T., Leysen, K., Hoogeman, M., Storme, G., Ridder, M.D.: Geometric accuracy of a novel gimbals based radiation therapy tumor tracking system. Radiother. Oncol. **98**(3), 365–372 (2011)
3. Lindl, B.L., Müller, R.G., Lang, S., Lablanca, M.D.H., Klöck, S.: TOPOS: a new topometric patient positioning and tracking system for radiation therapy based on structured white light. Med. Phys. **40**(4), 042701 (2013)
4. Price, G.J., Parkhurst, J.M., Sharrock, P.J., Moore, C.J.: Real-time optical measurement of the dynamic body surface for use in guided radiotherapy. Phys. Med. Biol. **57**(2), 415 (2012)
5. Rosten, E., Porter, R., Drummond, T.: Faster and better: a machine learning approach to corner detection. IEEE Trans. Pattern Anal. Mach. Intell. **32**(1), 105–119 (2010)
6. Schaerer, J., Fassi, A., Riboldi, M., Cerveri, P., Baroni, G., Sarrut, D.: Multi-dimensional respiratory motion tracking from markerless optical surface imaging based on deformable mesh registration. Phy. Med. Biol. **57**(2), 357 (2012)
7. Wasza, J., Bauer, S., Hornegger, J.: Real-time motion compensated patient positioning and non-rigid deformation estimation using 4-D shape priors. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part II. LNCS, vol. 7511, pp. 576–583. Springer, Heidelberg (2012)
8. Wasza, J., Bauer, S., Hornegger, J.: Real-time respiratory motion analysis using manifold ray casting of volumetrically fused multi-view range imaging. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013, Part II. LNCS, vol. 8150, pp. 116–123. Springer, Heidelberg (2013)

An Iterative Closest Point Framework for Ultrasound Calibration

Elvis C.S. Chen[1,2(✉)], A. Jonathan McLeod[2], John S.H. Baxter[2], and Terry M. Peters[1,2]

[1] Robarts Research Institute, Western University, London, Canada
[2] Biomedical Engineering Graduate Program, Western University, London, Canada
{chene,jmcleod,jbaxter,tpeters}@robarts.ca

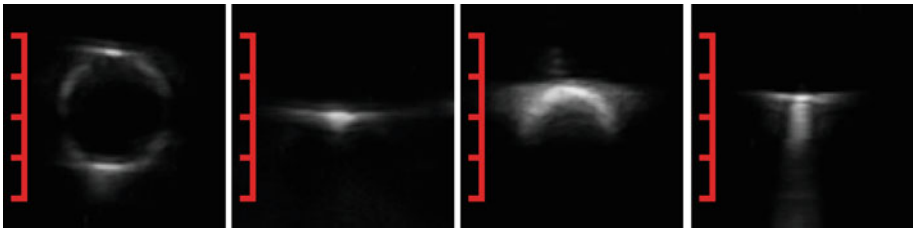
Abstract. We introduce an Iterative Closest Point framework for ultrasound calibration based on a hollow-line phantom. The main novelty of our approach is the application of a hollow-tube fiducial made from hyperechoic material, which allows for highly accurate fiducial localization via both manual and automatic segmentation. By reducing fiducial localization error, this framework is able to achieve sub-millimeter target registration error. The calibration phantom introduced can be manufactured inexpensively and precisely. Using a Monte Carlo approach, our calibration framework achieved 0.5 mm mean target registration error, with a standard deviation of 0.24 mm, using 12 or more tracked ultrasound images. This suggests that our framework is approaching the accuracy limit imposed by the tracking device used.

1 Introduction

Ultrasound (US) has become a pivotal component in many image-guided interventions, providing information support by visualizing anatomy beneath the organ surface. For example, US is used to monitor real-time needle advance in US-guided spinal anaesthesia [11] and is used routinely in minimally invasive abdominal and cardiac procedures. To aid spatial reasoning, a tracking system can be integrated into the surgical work flow, allowing the pose (orientation and position) of the US transducer and other tracked objects to be determined in a common coordinate frame. This enables advanced visualization such as image fusion between US and endoscopic video [7] and augmented virtuality environments for cardiac surgery [14]. For any navigated-US intervention, accurate US calibration is critical to the success of the procedure.

Ultrasound calibration [13] remains an active research topic. Categorically, one common means of achieving US calibration is to employ a tracked *line fiducial* as the calibration phantom. Muratore and Galloway [15] employed a calibrated pointer carefully placed in the US beam, producing a bright, sharp image of the pointer tip. The location of this feature, or fiducial, is determined both in the US image space and the tracker space, serving as the basis for solving the US calibration using a closed-form, least-squares, solution [10]. The accuracy of this method depends on how well the pointer tip can be placed on the ideal US

C.A. Linte et al. (Eds.): AE-CAI 2015, LNCS 9365, pp. 69–79, 2015.
DOI: 10.1007/978-3-319-24601-7_8



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

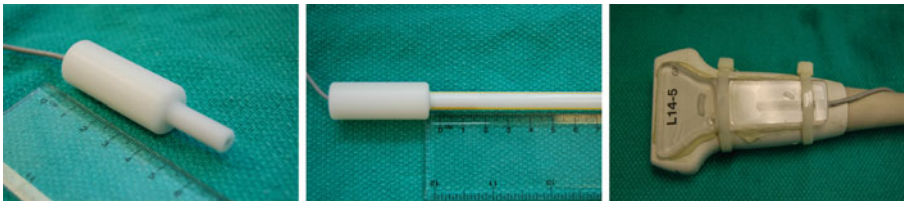
Fig. 1. Appearance of various line fiducials, scale: each tick mark = 2.5 mm: (a) hollow straw, outer diameter = 7.0 mm, (b) braided wire, diameter = 0.3 mm, (c) plastic sphere, diameter = 3.18 mm, and (d) 24 G spine needle, diameter = 0.51 mm.

mid-plane, and how well these fiducials can be localized in both the calibration phantom and the US image [17]. Khamene and Sauer [12] relaxed the requirement of the precise tip placement, by imaging the tool shaft instead, and using the orientation of the tracked line fiducial as the basis for US calibration. The mathematical formulation of their approach resolved the calibration parameter through numerical optimization, thus their method is sensitive to initialization and the configuration of the tracked poses.

Another common line fiducial phantom is the z-phantom [5,6,8], with wires or rods arranged in coplanar z patterns. The coplanar lines appear as colinear fiducials in the US image: by similar triangles, the location of the middle fiducial can be determined in the tracker space. The z-phantom eliminates the need for precise placement of the calibration phantom in the US beam, while allowing the closed-form and least-squares solution to be applied. Multiple z-fiducials can be placed within the phantom at varying depths and orientations, increasing the number of fiducials acquired per image while increasing robustness of the approach. However, the requirement for the z-phantom approach is that the phantom needs to be manufactured precisely, and the required image processing as the number of z-fiducials increases. Moreover, the configuration of the z-fiducials must be optimized for transducer geometry to accommodate US fan width and depth.

A variation of the Z-phantom was proposed by Welch et al. [18]. Instead of solving the calibration parameter using closed-form solution, they reformulated the calibration as a rigid body Iterative Closest Point [3] (ICP) point-to-line registration. Their phantom consisted of multiple wires stretched taut between two parallel plates. The configuration of the wire was not constrained in the z pattern. Correspondences between US fiducials and lines were assumed.

In all the line-fiducial based calibration methods, the centroid of the hyperechoic signal is used as the US fiducial [5,9,18]. Due to factors such as specular reflection, beam width, ringing effects, and low resolution outside of the US focal zone, these hyperechoic signal often appear wide and blurred in US image (Fig. 1). In our experience, the centroids of these US reflections are hard to identify accurately and robustly, resulting in large variation in the Fiducial



(a) (b) (c)

Fig. 2. (a) the concentric cylinder base of the straw phantom, with integrated magnetic sensor, (b) assembled straw phantom: common drinking straw provides a straight line fiducial, (c) a linear probe with magnetic sensor rigidly attached.

Localization Error (FLE). In contrast, a hollow-tube fiducial made with hyperechoic material can be easily segmented in B-mode images [1]. Based on this observation, we introduce an inexpensive calibration tool comprising of a single hollow straw. Because the assumption of scale isotropy in the US image may be incorrect (the scaling in the lateral direction being defined by the transducer element spacing, and the axial by the speed of sound in the medium), we further introduce a variation of the Iterative Closest Point (ICP) algorithm to determine the US calibration transformation incorporating anisotropic pixel scales as well as the rotation and translation.

2 Methods and Materials

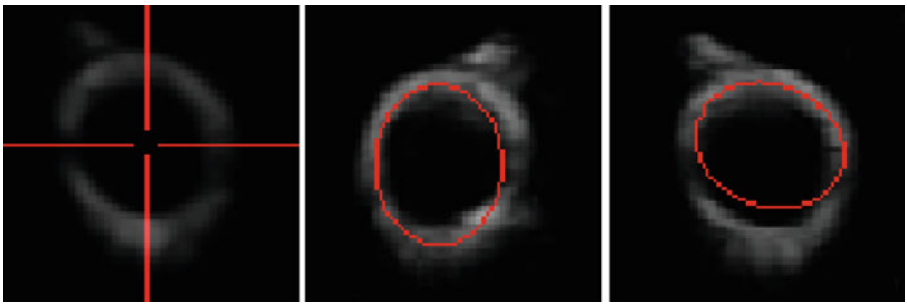
We introduce a "straw phantom", comprising a tracking sensor and a hollow straw, both attached to a concentric cylinder base (Fig. 2a). The base can be constructed precisely and inexpensively using a lathe. The axis of the cylindrical base, hence the direction of the straw, can be calibrated to the tracking sensor (Aurora, NDI, Canada) precisely by rotating it about a fixed axis.

A linear probe (L14-5/38, Ultrasonix, Canada) augmented with a 6-DOF magnetic (NDI, Canada) sensor was used (Fig. 2c). A standard US acquisition protocol was used: 10 MHz, 8 cm, with gain of 55 percent. The single focal point was set the the center of the image (4 cm depth). The digital US image were acquired using a frame-grabber (Epiphan System, Canada) at the size of 1280×1024 pixels.

2.1 Centroid Segmentation

The hollow cylindrical straw has low attenuation, producing either a circular or elliptical reflection in US image depending on the angle of insonation. The centroid of such a fiducial can be manually identified with the aid of a cross-hair cursor to a high degree of consistency (Fig. 3a).

For automatic segmentation, a thresholding technique [16] was first applied to binarize the US image. The edge of the fiducial was extracted by a morphological



(a) (b) (c)

Fig. 3. Segmentation of the hollow straw in US image: (a) manual centroid identification using a cross-hair cursor, (b) automatic ellipse fitting, and (c) automatic ellipse fitting with visible error.

closing operation on the resulting binary image to remove missing pixels, followed by a morphological opening operation to remove clutter. A randomized Hough transform was applied to fit an ellipse to the resulting edge and the centroid was used as the US fiducial (Fig. 3b,c).

2.2 Anisotropic Scaled Iterative Closest Point Algorithm

We formulate the calibration as a special case of the point-to-line ICP registration, where the correspondence between a single point (US fiducial) and a line (tracked straw) is known explicitly. Denote ${}^{C}T_{p}$ and ${}^{C}T_{l}$ as the tracked pose of the tracker sensor attached to the probe (p) and line fiducial (l), respectively, and C the common coordinate system (i.e. tracker). Let ${}^{us}f = (f_x, f_y, 0, 1)$ be the pixel location of a fiducial, and the paired points $({}^{l}a, {}^{l}b)$ denote the end points of the straw phantom. Applying tracked transformations to the fiducial,

$$ {}^{l}f = ({}^{C}T_{l})^{-1}({}^{C}T_{p})({}^{p}T_{us})^{us}f \tag{1} $$

describes the location of f in l, where it must lie within the line segment $\overline{{}^{l}a{}^{l}b}$. The similarity matrix ${}^{p}T_{us}$ is the ultrasound calibration matrix we seek. At each iteration, the point q on line $\overline{ab}$ closest to f is used as the corresponding point.

Ordinary ICP does not solve for a similarity transform although isotropic scaling can be trivially incorporated into the underlying transformation [10]. The error metric to minimize for ICP is the Fiducial Registration Error (FRE):

$$ FRE^2 = \frac{1}{N_p}\sum_{i=1}^{N_p} |(RS({}^{us}p_i) + t) - ({}^{p}q_i)|^2 \tag{2} $$

where R is the orthonormal rotation, $S = diag(s_x, s_y, s_z)$ the diagonal scaling, t the translation, and N_p is the number of fiducials.

We [4] recently introduced a generalization of ICP to solve for anisotropic scaling between datasets, where the underlying solution for the Anisotropic-Scaled Orthogonal Procrustes Problem (ASOPP) was based on the recent theoretical development by Bennani Dosse and Ten Berge [2]. Given two datasets X and Y with one-to-one correspondence, the pseudo-code for solving the registration between these two datasets involving anisotropic-scales is presented below:

demean the data to obtain $\hat{X}$ and $\hat{Y}$;
normalize the row of $\hat{X}$ such that$\hat{X}_i \cdot \hat{X}_i{}^T = 1$;
while Δ *FRE* > *threshold* **do**
  $[U, \Lambda, V] = svd(B * diag(diag(R^T * B)))$;
  $R = U\ diag([1, 1, det(U * V)])\ V^T$;
  compute new FRE
end
$S = diag(B^T * R)$

As with the ordinary ICP, our Anisotropic Scaled ICP (ASICP) has monotonic convergence. To apply ASICP to solve US calibration, X_i is the pixel locations of the segmented US centroid, and Y_i is the point on tracked line-fiducial closest to X_i at each ASOPP iteration. Since there is only one line-fiducial, and hence one US fiducial per data collection, the correspondence between X_i and Y_i is explicitly known.

2.3 Extension to Bi-Plane Ultrasound

Modern US transducers employing 2D matrix-array have the ability to acquire simultaneous bi-plane images with a common origin, where the angles between each plane can be manually adjusted. Our proposed ICP algorithm naturally extends to this image modality, where 2 fiducials can be collected simultaneously (one from each plane) to further constrain the calibration transform.

For instance, when the bi-plane US images are mutually perpendicular, the fiducial at one image plane can be assigned a coordinate of $p1 = (p_{x1}, p_{y1}, 0)$, and $p2 = (0, p_{y2}, p_{z2})$ on the other (Fig. 4). In this mode of operation, the straw phantom intersects both planes at an oblique angle, allowing two US fiducials to be collected simultaneously with a single set of tracking data.

3 Results

We evaluate the proposed straw phantom using three variations of the ICP algorithm: the proposed anisotropic-scaled ICP (ASICP), isotropic-scaled ICP (Iso-ICP), and pre-scaled rigid ICP (Pre-ICP). For rigid ICP, the isotropic scale was estimated using the on-screen caliper, and applied to the pixel coordinate system prior to calibration (Fig. 5).

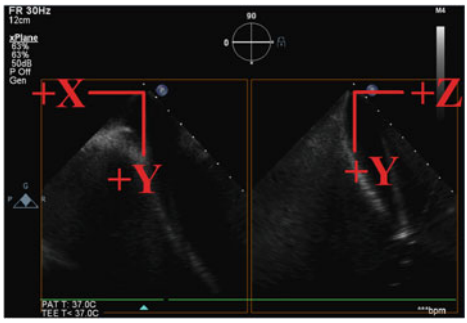


Fig. 4. Modern US scanners has the ability to acquire simultaneous bi-plane US images. When they are mutually perpendicular to each other, a common coordinate system can be assigned in which both US images share a common origin and an axis.

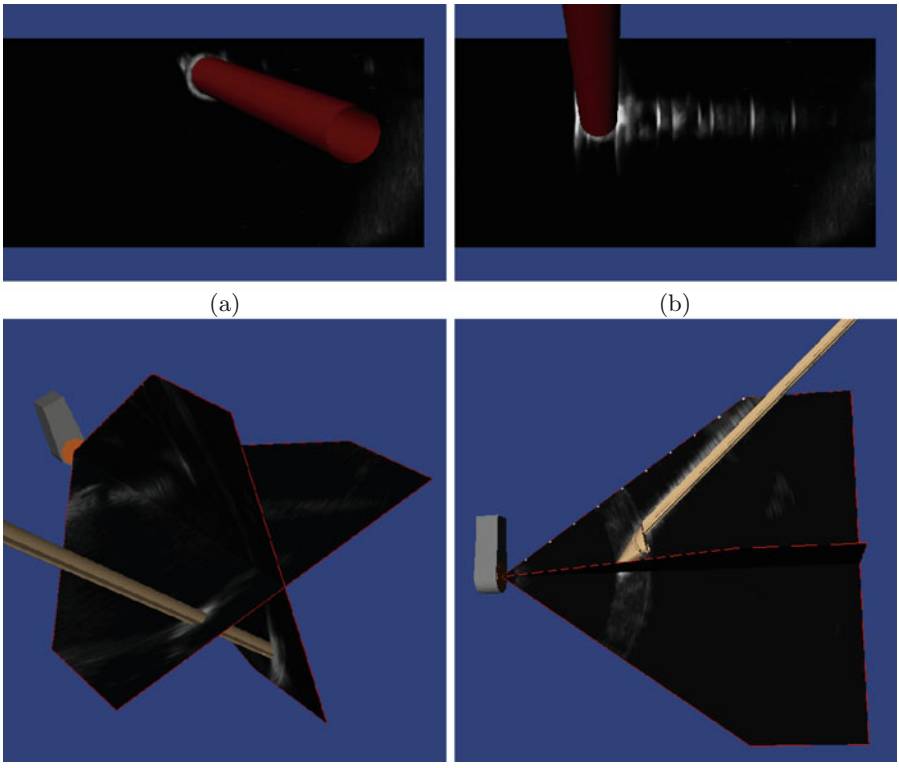


Fig. 5. Visualization of the proposed US calibration: (a,b) straw phantom intersecting with the US beam, producing an elliptical hyperechoic reflection, (c,d) extension of the US calibration to bi-plane US. Note the tip and orientation of the surgical tools aligns well with US reflection.

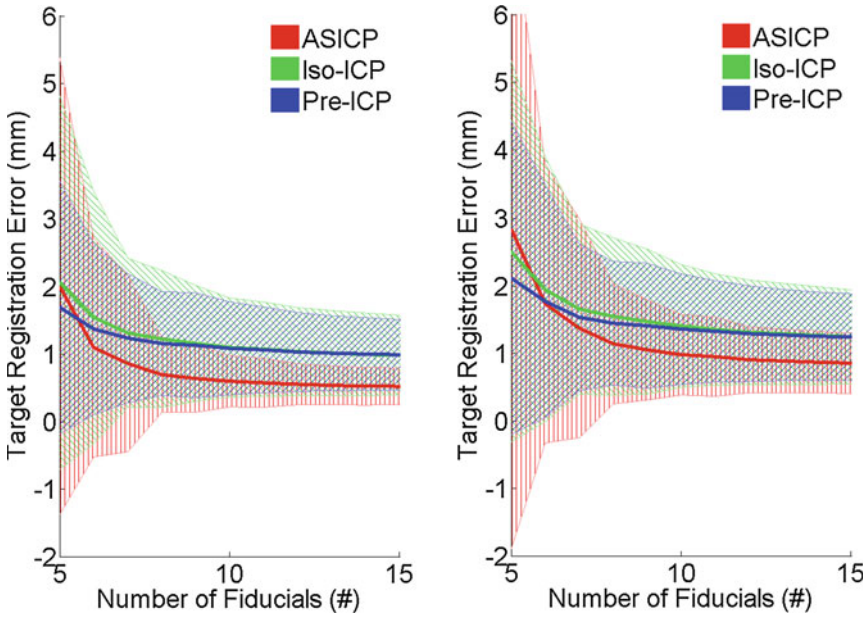


Fig. 6. Mean and standard deviation of absolute TRE using (a) manual segmentation, (b) automatic segmentation. The TRE is measured as the error between the centroid and the line fiducial projected onto the ultrasound image.Mean and standard deviation of absolute TRE using (a) manual segmentation, (b) automatic segmentation. The TRE is measured as the error between the centroid and the line fiducial projected onto the ultrasound image.

A total of 100 datasets were acquired, each comprising the US image and the tracked poses of the straw phantom and US transducer. Both manual and automatic identification of the fiducial centroids were obtained. The straw phantom was moved uniformly throughout the imaging volume.

A Monte Carlo approach was employed to assess the robustness and accuracy of the proposed US calibration algorithm. A set of k datasets were randomly sampled from the total of 100 ($k \in [5..40]$), based on which a set of 6 calibrations were derived (3 ICPs $\times$ 2 segmentations) For each calibration, the remaining $100 - k$ datasets were treated as targets to calculate TRE. For each k, this experiment was repeated 1000 times.

Figure 6 displays the spread of the TRE for $k \in [5..15]$ in terms of mean (solid line) plus or minus one standard deviation (hatched area). Using both manual and automated segmentation, there is an inversion in terms of the preference of calibration technique. At low numbers of fiducials (< 7), both isotropic and prescaled ICP outperform ASICP, largely due to overfitting based on the number of degrees-of-freedom in the calibration transform.

As the number of fiducials used to calibrate becomes larger, the error in each transform decreases as does the standard deviation of said error, reaching a horizontal asymptote dependent on the tracking accuracy and FLE. The difference between automatic and manual segmentation can give an estimate as to how much of this asymptote is attributable to segmentation error (Table 1).

Table 1. RMS, mean, and std of the mean TRE, k = 40

mTRE (mm)	ani auto	ani manu	iso auto	iso manu	pre auto	pre manu
rms	0.86	0.52	1.27	1.02	1.27	1.02
mean	0.77	0.47	1.14	0.90	1.16	0.92
std	0.38	0.23	0.57	0.47	0.53	0.42

For all variants of ICP, calibration based on manual segmentation works significantly better, suggesting that our automatic ellipse fitting algorithm can benefit from further improvement. Despite the error in fiducial localization, all variants of the ICP calibrations achieved a mean-TRE of less than 1.2 mm, when more than 20 fiducials were used to derived the calibration. For all ICP using manual segmentation, sub-millimeter mean-TRE can be achieved with only 12 images. ASICP performed the best, achieved a mean-TRE of 0.5 mm when using more than 12 fiducials (Fig. 7).

To validate the scale, we constructed a grid phantom with 4 parallel thin wires. The US transducer was carefully placed on top of the grid phantom, ensuring all 4 fiducials were aligned in both vertical and horizontal direction. The US probe was then rotated axially until the US fiducials are at the brightest, at when the US image was acquired. The top-center edge of each fiducial was carefully identified, and scales calculated serving as ground truth (Table 2).

ASICP successfully compensated for anisotropic scaling. In the axial direction, the scaling factor was reduced by 3.5 % relative to the on-screen caliper. This corresponds to the 3.8 % reduction in speed of sound from the assumed 1540 m/s to that of the room temperature water bath. Additionally, the scaling factor in the lateral direction was slightly higher than in the axial direction. This was due to the aspect ratio of the display during screen capture and represents another source of anisotropic scaling that is often unaccounted for. Both the axial and lateral scale factors found by ASICP agreed closely with the ground truth measurements obtained using the grid phantom. On the other hand, isotropic ICP produced a single scaling factor between those found for the axial and lateral dimensions resulting in less accurate calibration.

ASICP consistently over estimates the scaling in the axial direction to reflect the decrease in speed of sound. A consequence of which is it under-estimates the scaling in the lateral directions. As a reference, the pixel scaled determined using the on-screen caliber is 0.125 mm per pixel. Isotropic ICP over estimates the scaling as well to reflect the speed of sound

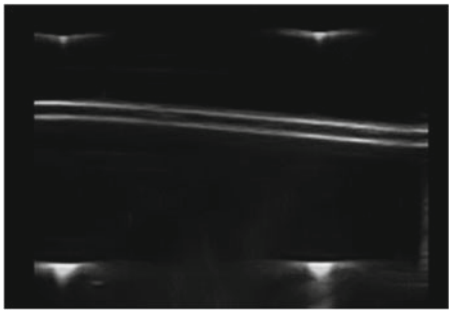


Fig. 7. US image of the wire-grid phantom: the physical distance between wires in lateral and axial directions are 25 mm and 20 mm, respectively. The pixel coordinate for each fiducial are: $(468, 215), (652, 215), (468, 380), (652, 380)$.

Table 2. Scales determined by ICP algorithms.

Scales (mm/pixel)	Iso-ICP auto/manu	ASICP auto/manu	Ground Truth
Lateral	0.1273/0.1268	0.1369/0.1357	0.1359
Axial		0.1206/0.1208	0.1212

4 Conclusion

We propose an Anisotropic Scaled Iterative Closest point framework for solving ultrasound calibration, where the transformation we seek is the anisotropic scales, followed by rotation and translation. We also introduce a readily constructed hollow-tube phantom which is easier to segment from ultrasound image. By reducing the Fiducial Localization Error through accurate segmentation, our framework is able to achieve sub-millimeter accuracy with as little as 12 data acquisitions. An automatic segmentation algorithm for this phantom is proposed which greatly reduces the labor spent in ultrasound calibration.

Anisotropic scaled ICP provides a physically complete model of the transformation required for ultrasound calibration. This model resulted in lower mean and standard deviation of absolute TRE, approaching the fundamental limit of the tracking system. ASICP also accurately estimates the physical size of the pixels in both the lateral and axial dimensions. Relatively few additional fiducials are necessary to prevent overfitting of the calibration transform.

The spatial configuration of the line fiducials provides a direct and physical constrain for the ultrasound calibration parameters. Consider, as an example, all tracked line-fiducials are parallel to each other. Such spatial configuration would result in an infinite number of ultrasound calibrations as translation along parallel line-fiducials cannot be constrained. For future work, we intent to work on a theoretical model for optimizing line-fiducial configuration [19], incorporate anisotropic uncertainties into calibration method, and investigate automatic *in situ* calibration in which the speed of sound in human tissue is intraoperatively compensated for.

References

1. Ameri, G., McLeod, A.J., Baxter, J.S.H., Chen, E.C.S., Peters, T.M.: Line fiducial material and thickness considerations for ultrasound calibration. In: Proceedings of SPIE, vol. 9415, pp. 941529–941529-9 (2015)
2. Bennani Dosse, M., Ten Berge, J.: Anisotropic orthogonal procrustes analysis. J. Classif. **27**, 111–128 (2010)
3. Besl, P., McKay, N.D.: A method for registration of 3-d shapes. IEEE Trans. Pattern Anal. Mach. Intell. **14**(2), 239–256 (1992)
4. Chen, E.C., McLeod, A.J., Baxter, J.S., Peters, T.M.: Registration of 3D shapes under anisotropic scaling. Int. J. Comput. Assist. Radiol. Surg. **10**(6), 867–878 (2015)
5. Chen, T.K., Thurston, A.D., Ellis, R.E., Abolmaesumi, P.: A real-time freehand ultrasound calibration system with automatic accuracy feedback and control. Ultrasound Med. Biol. **35**(1), 79–93 (2009)
6. Cheng, A., Ackerman, M.K., Chirikjian, G.S., Boctor, E.M.: Design and development of an ultrasound calibration phantom and system. In: Proceedings of SPIE, vol. 9036, pp. 903624–903624-8 (2014)
7. Cheung, C.L., Wedlake, C., Moore, J., Pautler, S.E., Peters, T.M.: Fused video and ultrasound images for minimally invasive partial nephrectomy: a phantom study. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part III. LNCS, vol. 6363, pp. 408–415. Springer, Heidelberg (2010)
8. Comeau, R.M., Fenster, A., Peters, T.M.: Integrated MR and ultrasound imaging for improved image guidance in neurosurgery. In: Proceedings of SPIE, vol. 3338, pp. 747–754 (1998)
9. Gobbi, D.G., Comeau, R.M., Peters, T.M.: Ultrasound probe tracking for real-time ultrasound/mri overlay and visualization of brain shift. In: Taylor, C., Colchester, A. (eds.) MICCAI 1999. LNCS, vol. 1679, pp. 920–927. Springer, Heidelberg (1999)
10. Horn, B.K.: Closed-form solution of absolute orientation using unit quaternions. J. Opt. Soci. Am. A **4**(4), 629–642 (1987)
11. Karmakar, M.K., Li, X., Ho, A.M.H., Kwok, W.H., Chui, P.T.: Real-time ultrasound-guided paramedian epidural access: evaluation of a novel in-plane technique. Br. J. Anaesth. **102**(6), 845–854 (2009)
12. Khamene, A., Sauer, F.: A novel phantom-less spatial and temporal ultrasound calibration method. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3750, pp. 65–72. Springer, Heidelberg (2005)
13. Mercier, L., Langø, T., Lindseth, F., Collins, L.D.: A review of calibration techniques for freehand 3-D ultrasound systems. Ultrasound Med. Biol. **31**(2), 143–165 (2005)
14. Moore, J., Chu, M., Kiaii, B., Bainbridge, D., Guiraudon, G., Wedlake, C., Currie, M., Rajchl, M., Patel, R., Peters, T.: A navigation platform for guidance of beating heart transapical mitral valve repair. IEEE Trans. Biomed. Eng. **60**(4), 1034–1040 (2013)
15. Muratore, D.M., Galloway Jr., R.L.: Beam calibration without a phantom for creating a 3-D freehand ultrasound system. Ultrasound Med. Biol. **27**(11), 1557–1566 (2001)
16. Otsu, N.: A threshold selection method from gray-level histograms. IEEE Trans. Syst. Man Cybern. **9**(1), 62–66 (1979)
17. Prager, R., Rohling, R., Gee, A., Berman, L.: Rapid calibration for 3-D freehand ultrasound. Ultrasound Med. Biol. **24**(6), 855–869 (1998)

18. Welch, J., Bax, M., Mori, K., Krummel, T., Shahidi, R., Maurer, C.: A fast and accurate method of ultrasound probe calibration for image-guided surgery. In: Proceedings of Computer Assisted Radiology and Surgery, p. 1078 (2002)
19. West, J.B., Maurer, C.R.J.: Designing optically tracked instruments for image-guided surgery. IEEE Trans. Med. Imaging **23**(5), 533–545 (2004)

Development of 4D Human Body Model that Enables Deformation of Skin, Organ and Blood Vessel According to Dynamic Change

Naoki Suzuki[1(✉)], Asaki Hattori[1], and Makoto Hashizume[2]

[1] Institute for High Dimensional Medical Imaging,
The Jikei University School of Medicine, Tokyo, Japan
{nsuzuki,hat}@jikei.ac.jp
[2] Department of Advanced Medical Initiatives, Faculty of Medical Sciences,
Kyushu University, Fukuoka, Japan
mhashi@med.kyushu-u.ac.jp

Abstract. Any part of a live human body bears constant dynamic changes in the spatial and temporal axis. Therefore we believe that analyzing and understanding the 4 dimensional dynamics of a human body will contribute to new inputs in medical diagnosis and doctor's judgment for effective treatment. To achieve this objective, we are developing a quantitative 4D human body model with inner structures such as the skeletal structure and major organs. We are aiming to grasp from various viewpoints, the spatiotemporal (four-dimensional) changes of an anatomical structure of a live human being. The model in this research is constructed based on a subject measured by MRI. The aim is to have the model's inner structures change according to the subject's full body movement data. In addition, we aim to have a function in which the shape of the skin surface changes synchronizing with the data.

In this research, we examine the possibilities of clinical application of our 4D model that not only has skeletal and major organ blood vessel systems but also has muscular systems as inner structure.

Keywords: 4D human model · Elastic skin surface · Motion analysis

1 Introduction

It is still difficult with current diagnostic imaging methods to measure the dynamic changes of the whole human body as 4D phenomenon and analyze them in quantitative way. It is true that we are able to carry it out in separate parts of the body, such as measuring the heart dynamics by cardiac MRI and measuring lung and digestive tract by MDCT. But currently, there are no diagnostic devices that measure the 4D changes of the inner body when the whole body moves, for example when humans walk, stand or sit down.

Therefore, we need a method that can observe and analyze the human dynamics in 4D so that we have the comprehensive understanding of the human body and the quantitative grasp of the medical condition from various aspects.

C.A. Linte et al. (Eds.): AE-CAI 2015, LNCS 9365, pp. 80–91, 2015.
DOI: 10.1007/978-3-319-24601-7_9

We believe that a 4D human body model that can analyze the dynamic changes without restraint and to understand quantitatively the spatio-temporal (four-dimensional) changes of the anatomical structure of the human body will lead to improving future medical practice. We therefore endeavored to manipulate a human body model that has an inner structure including skeletons, major organs, blood vessel and muscular systems and manipulate it in real time so that we can observe and analyze it without restraint.

2 Methods

2.1 Construction of 4D Drive Model

The 4D human model mentioned here is not a model like a sculpture designed by software and a creator. Rather, it is a quantitative structure model system that is reconstructed by extracting anatomical shape from the subject's oriented data and has a function that can drive the model according to dynamic data derived from the subject.

There are two major steps to drive this human model. One is to construct a skeletal model that can show body dynamics by moving the body model's "non-deformable skeletal structure" around each joint based on the position data of optical markers spread throughout the body surface.

The second process is to generate dynamic changes in the soft tissue system such as internal organ systems, blood vessel systems and muscular systems that deform their shapes and positions according to the dynamic changes in the "non-deformable skeletal structure".

There are already various researches on the former, i.e. the analysis of the dynamic changes of the skeletal system [1–6]. We have conducted clinical cases where we assume the movements of the skeletons using body surface motion capture data and applying them to surgery planning, surgery assistance and after surgery management [7–9].

2.2 Generation Process Void of "Anatomical Contradiction"

The most important thing in constructing a 4D human body model is to generate proper deformation in the soft tissue system surrounding the "non-deformable skeletal structure". To realize this, we need to create a deformation rule similar to that of natural soft tissue deformation. In addition, this calculation process for this rule needs to be fast enough to deform the data volume of the whole human body in real time to maintain the effectiveness of the model. Therefore, the rule must enable the soft tissues to deform in an anatomically reasonable manner without having a significant burden on the calculation inside the workstation. There must be a soft tissue deformation rule that drives the bodies of humans and other animals effectively. It is a known fact anatomically, that soft tissues of organisms surrounding the skeletal system do not hugely change their positions nor do their layered structure invert in normal body motions. This non-contradictory deformation corresponding to organ deformation also occurs in the vascular system including the blood vessel system in the organs that are spread in a dendritic structure in the body.

We think that ultimately, the skin that layers outside these contents in the body also deform according to the same deformation rule of "void of anatomical contradiction". We mention below 6 points to the deformation rule.

(1) Create "Branch off Dendritic Model" structure where the non-deformable skeletal system changes directions with each joint at their centers. Create a rule where the surrounding soft tissues systematically deform according to this structure.

(2) Generate subordination relation between dendritic model that has multiple joints and their surrounding soft tissue structures. Carry out movement and deformation based on skeletal infliction and movement.

(3) Decide the deformation volume using the distance between each soft tissues and the nearest dendritic model as parameter. Generate relative deformation that is void of "anatomical contradiction" by generating serial deformation and movement.

(4) Do not change the attachment position on the skeletal model of the skeletal muscle on the muscle model. The muscle deformation will be restricted by the movement of the attachment position that follows the movement of the skeletal system along with the deformation mentioned above.

(5) The skin surface shape will deform in the same process as mentioned above. If the deformation occurs that does not contradict with anatomy, abnormal deformation such as the inner structure projecting from the skin that is farthest from the dendritic model should not occur based on this rule.

(6) If we set the parameter based only on the distance to the nearest dendritic model for branched off dendritic model, it will generate abnormal deformation called "deformation accident". This is because if an anatomically non-subordinate dendritic model is situated near the branched off model, it will be affected. We will sectionalize body parts and calculate the deformation volume of each section to prevent this, so that we will be able to restrain abnormal deformation that will be affected by other parts that approach it due to body movement.

We used linear blend skinning, a method commonly used in computer graphics to deform soft tissues.

Regarding Point **P** on the model before deformation, Point **P(t)** after deformation at a point in time t can mathematically be expressed as below if there are n number of joints that influence **P** and there are **Mi(t)** changes from the basic posture of each joint.

$$\begin{aligned} \mathbf{P}(\mathbf{t}) &= \sum_{i=1}^{n} w_i \mathbf{M_i}(t)\mathbf{P} \\ \sum_{i=1}^{n} w_i &= 1, w_i \geq 0 \end{aligned} \tag{1}$$

Weight w_i, is the sum of distance d_i which is the distance between **P** and joint j_i, and distance d_{i+1} which is the distance between **P** and adjacent joint j_{i+1}.

$$w_i = (d_i + d_{i+1})^{-c} \tag{2}$$

Currently, we are using constant c = 15 which is by experience known to be close to the deformation of human beings. But in the future, we feel the need to reflect the changes of joint peripherals and amount of changes at each body part that are even

closer to real human changes. We plan to carry this out by measuring the amount of changes in the soft tissues when each joint is in motion and reflecting this to the weight, so that we can change the weight according to body parts, sex and physical size.

In addition, we use GPU for deformation calculation to realize deformation of the inner structure in real time. We can process the whole body movement such as walking in approximately 40fps. We did not use commercially available software and developed the system on our own as we plan to use it for human inner structure real time analysis in the future.

2.3 Data Collection for Creating Whole Body Model

The whole human body structure was measure by MRI (Slice thickness: 2 mm by T1 weighted image, slice amount: 840 slices). We took the shapes of body surface, main organs, skeletal system, and vascular system from MRI data and segmented them using Analyze software (Mayo Clinic, USA) and made them surface models. We also segmented major muscle shape for the skeletal muscular system and designated attachment position for each muscle in the skeletal framework. We set the joints so that the skeletal system could be bent and rotated surrounding each joint.

We obtained the heart drive from an MRI cluster (20 slices of coronary tomographical images) taken under different condition from the rest of the body. We constructed a serial 3D image cluster of a heartbeat divided into 15 phases. We segmented the outer cardiac wall, right and left atriums and ventricle lumen from this and created a 4D heart cardiac model.

We used surface body model (628,940 polygon), abdominal organ 31 model (1,074,432 polygon), vascular system 71 model (394,250 polygon), lower limb skeletal muscle 70 model (73,252 polygon) and skeleton 223 model (942,220 polygon).

2.4 Motion Measurement

We used a 3D motion analysis system VICON512 (Vicon Motion Systems, UK) and measured the motions of the head, body trunk, right and left upper and lower leg with 42 optical markers. We converted the dendritic model (Fig. 1 green model) that was constructed from the position data of the optical markers to a dendritic model (Fig. 1 blue model) that corresponds to the anatomical shape of the skeletal model and joint cluster. We ultimately created a 4D model (Fig. 1 front row model) generating deformation of major organs, blood vessel system, skeletal muscular system and skin.

2.5 Verification of Method

In verifying the active changes of each part of the whole human body of the model we have developed this time, i.e. the flexing motions accompanying human actions, it is difficult to measure and examine the changes in the inner structure of a human body with non-invasive way. Therefore, we examined the accuracy of the model by

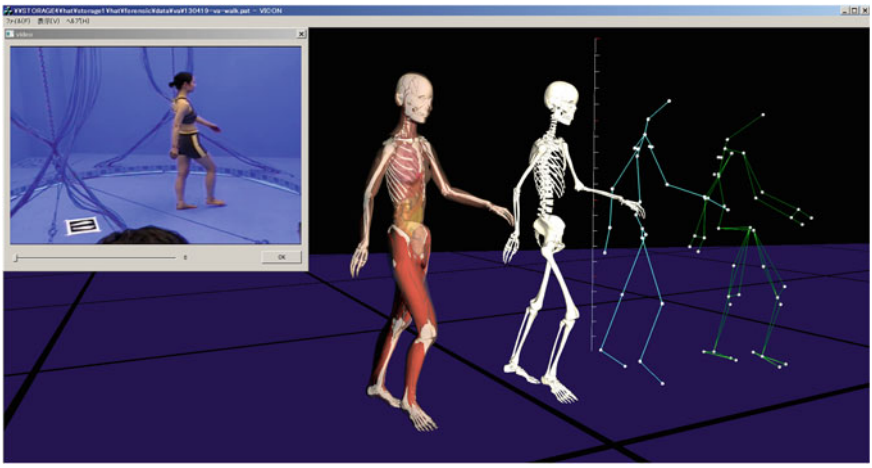


Fig. 1. Video image of subject being measured in walking motion (upper left) and manipulation example of 4D model (from far right: dendritic model by optical marker, dendritic model corresponding to anatomical shape of skeletal model and joint cluster, skeletal whole human body model, whole human body model including major soft tissues) (Color figure online).

measuring the body surface shape, specifically the changes in elasticity of the skin by using optical markers.

We positioned optical markers near joints, where the skin largely changes in elasticity through body motion and measured the changes in optical markers when the body is in motion using VICON512. We targeted upper and lower limb region. We set up 3 markers in the targeted region and measured the position of each marker during flexing motion. Next, we calculated the elasticity from the changes of skin from the changes of the triangle consisted by the 3 points. We also set up 3 measuring points on the skin of upper and lower limbs of the 4D whole human body model and calculated the shape changes of the triangle according to the flexing motions. We compared the two.

The first marker in the upper limb is situated at the olecranon and consists the top of the triangle. The second marker is situated near the armpit, which overlaps with the inner border of the deltoid muscle. This consists one point in the base of the triangle. The third marker is set up to make up an isosceles triangle with the other two points. It is set up at the outer border of the deltoid muscle. The measurement was done in a situation where the body did not move the upper arm and flexed the elbow to the periphery (Fig. 2).

The first marker in the lower limb is situated in the upper border of the patella and consists the top of the triangle. The second maker is situated at the center of the vastus lateralis muscle in the middle level of the upper border of the patella and the anterior superior iliac spine. We chose the inner border of the vastus lateralis muscle for the third marker to make up an isosceles triangle with the other two points. The measurement was done in a situation where the body pulled up the thigh to the horizontal position against the body (Fig. 3).

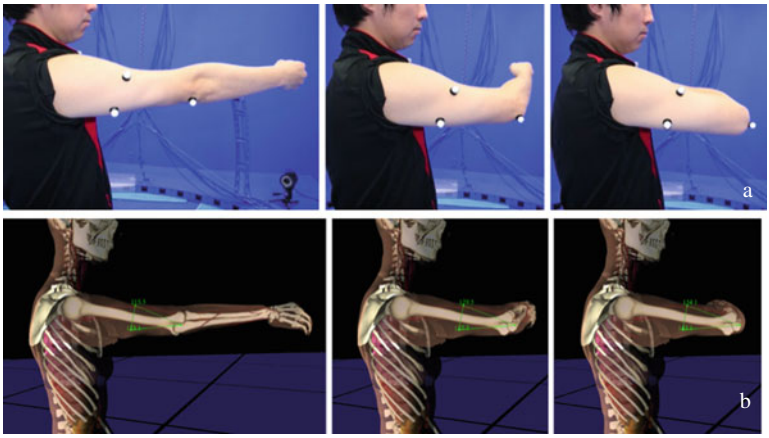


Fig. 2. Measurement of skin elasticity of upper arm during flexing of elbow joint (a: bending elbow by subject, b: bending elbow by 4D body model)

Fig. 3. Measurement of skin elasticity of thigh during flexing of knee joint (a: bending knee by subject, b: bending knee by 4D body model)

We measured the changes in the skin in the joint region by setting the markers at the elbow and the upper border of the patella, which consist the tops of the each triangle, as reference marks. We measured the distance of the other markers to the reference marks.

In the current verification process, we took the following method to match the measurement point on the subject's body surface and that of the 4D model. It is easy to specify the olecranon of the upper limb and the superior boarders of patella in the lower limb in both the subject and the model. But the 2 others do not have external form characteristics that are easy to specify and therefore we used subcutaneous muscle form as specification factors. A specialist located subcutaneous muscle in the subject body and we used translucent display to observe the muscle form on the model's skin model.

3 Results

3.1 4D Model Visualization

We show in the upper left of Fig. 1, the operational status of the whole body model from data of subject motions captured by VICON system. The process of corresponding optical marker information to the skeletal model and ultimately deforming soft tissues to display them as a whole body image covered with skin is as already explained in 2-4 Motion Measurement.

Figure 4 shows the status of deformation of the organs and blood vessel system without "anatomical contradiction" in all parts of the body by displaying the 4D model from the front (Fig. 4a) and right (Fig. 4a) in the form of time sequential images. This image shows the deformation of lower limb muscular system, vascular system including the beating heart and the inner structure that has major organs such lung, kidney, abdomen, and digestive system according to body motion. In addition, it displays the deformation of the skin that wraps these body contents.

Figure 5 shows deformation of abdominal region (Fig. 5a) and lower limb muscle (Fig. 5b). The body surface is transparent in order to clearly see the motions of the abdominal organs and lower limb muscular cluster when the body is walking. We could observe the motions of the lower muscular limb and we could also focus on a particular muscle and see its flexibility if needed.

3.2 Verification of the Model

There were 5 subjects and the experiment was conducted with total of 20 lower and upper limbs. We show in Fig. 6 examples of the changes in the 3 makers during measurement of subjects. In the figure, we show a case in which the distance between the reference mark and the other two markers is the shortest and the longest. The elasticity of the each joint region during flexing and bending motions resulted in an average of 22.8 ± 7.1 % for elbow joints and 18.9 ± 7.7 % for knee joints. For the same

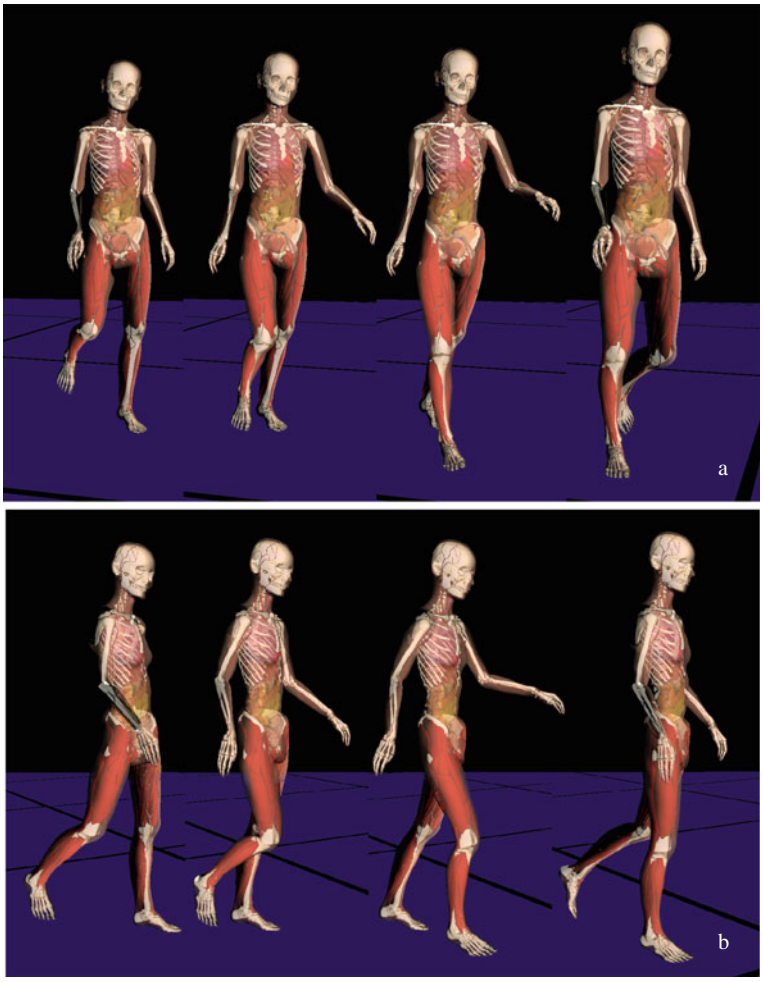


Fig. 4. Result of chronological display of 4D human body model manipulated by walking motion data a: viewed from front b: viewed from right

motions by the 4D whole body model, the result was 15.9 ± 3.8 % for elbow joint and 8.3 ± 3.0 % for knee joint (Fig. 7).

From the examination of the measurement of changes in the body surface shape during flexing motion, we admitted that the 4D human body model and the real human body showed approximately the same changes. Therefore we concluded that the model could be applied to clinical situations such the visualization of the changes in the human inner structure and diagnosis.

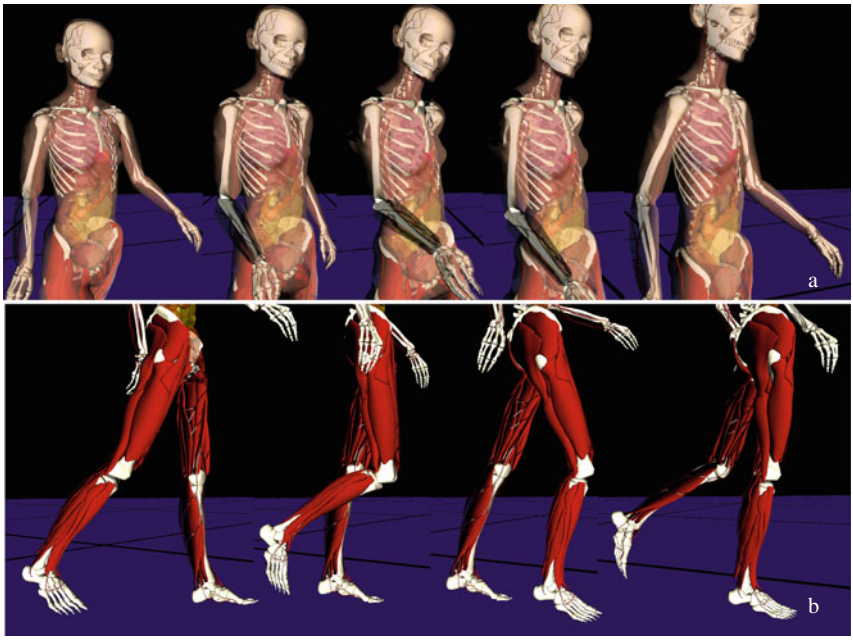


Fig. 5. Result of chronological display of abdominal region (a) and lower limb muscle (b) motion during walking

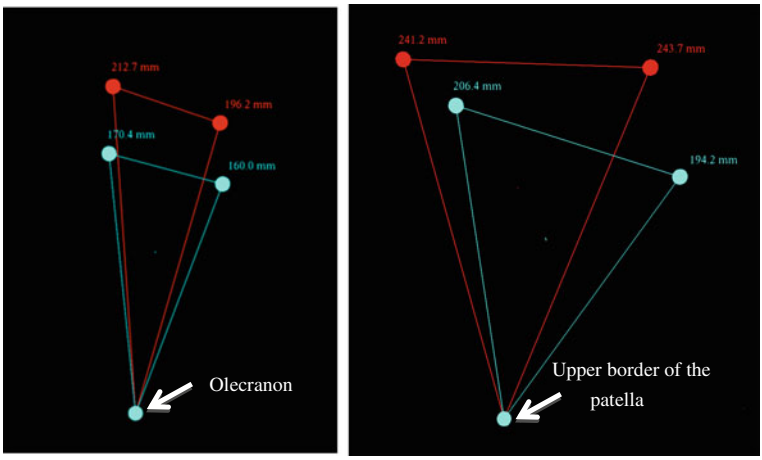


Fig. 6. Changes in the position of marker on the skin during subject's flexing motion. Extension is displayed in blue, bending is displayed in red. (a: bending of elbow joint, b: bending of knee joint) (Color figure online).

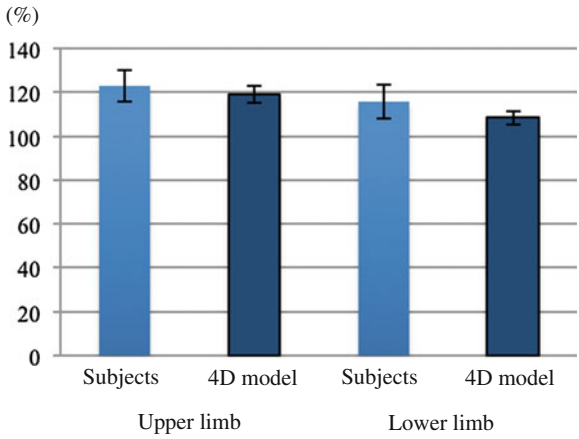


Fig. 7. Elasticity ratio of subjects and 4D model. Light blue shows subjects. Dark blue shows 4D model (Color figure online).

4 Conclusion

This 4D display has a function in which we can observe and measure the change according to body motion in real time. In addition by setting temporal axis randomly, it can observe both fast and slow movements.

With the method, we were able to quantitatively visualize changes in soft tissues and skin surface corresponding to human body structure movements by using patient oriented data, which are not possible to measure by current imaging diagnostic devices. Therefore we can say that we were able to construct a human model that is able to apply to various clinical situations. By using this method, we were able to display by image what surgeons and clinicians could only picture in their heads.

The mutual changes made in the skeleton and its surrounding skeletal muscles have been visualized and used in kinematic analysis [10–13]. But there are still very few methods to visualize the movements and the changes that occur from these movements in the field of organs and blood vessel systems that make up the trunk of the body using the patient's inner structure model and his/her motion data.

If this is realized, it will enable us the following in clinical medicine. (1) comprehend the location and changes made to the target organ when a patient takes complex posture on the surgical bed, (2) assume the changes of hemodynamics following posture changes, (3) comprehend the amount and direction of outside force applied to thoracic and abdominal organs. In addition, it will enable us the following in sports medicine and biotechnology. (4) Calculation of acceleration applied to organs when body is in motion, (5) calculation of kinetic energy for each body part, (6) shift in center of gravity during body motion, (7) understanding of energy consumption. And it will enable us the following in forensic medicine. (8) detailed inspection of how the victim's body was injured by outside force taking into account victim's body motion.

In the future, to evaluate the deformation of soft tissues in the deeper parts of the body, we plan to acquire as 4D data set, the deformation situation of a specific body part when a body carries out certain action using MRI, and compare them with the deformation situation on the same body parts on the model to conduct accuracy verification.

Going forward, we plan to examine the accuracy of outer deformation result by obtaining the deformation status of a particular body part corresponding to a motion using MRI as 4Ddata set for inner soft structure and examine the accuracy by comparing it with the same body part on the model. Based on the examination result, we plan to create a 4D human model that can generate more accurate deformation to approach those being generated in the real body structure. We also plan to develop a method to display the whole body as "volume image" in 4D by using this same deformation method as we see adopting the whole body volume data is effective in clinical situations.

References

1. Söderkvist, I., Wedin, P.Å.: Determining the movements of the skeleton using well-configured markers. J. Biomech. **26**(12), 1473–1477 (1993)
2. Schmidt, R., Disselhorst-Klug, C., Silny, J., Rau, G.A.: Marker-based measurement procedure for unconstrained wrist and elbow motions. J. Biomech. **32**(6), 615–621 (1999)
3. Rab, G., Petuskey, K., Bagley, A.: A method for determination of upper extremity kinematics. Gait Posture **15**(2), 113–119 (2002)
4. Cappozzo, A., Della Croce, U., Leardini, A., Chiari, L.: Human movement analysis using stereophotogrammetry: Part 1: theoretical background. Gait Posture **21**(2), 186–196 (2005)
5. Delp, S.L., Anderson, F.C., Arnold, A.S., Loan, P., Habib, A., John, C.T., Thelen, D.G.: OpenSim: open-source software to create and analyze dynamic simulations of movement. IEEE Trans. Biomed. Eng. **54**(11), 1940–1950 (2007)
6. van den Bogert, A.J., Geijtenbeek, T., Even-Zohar, O., Steenbrink, F., Hardin, E.C.: A real-time system for biomechanical analysis of human movement and muscle function. Med. Biol. Eng. Comput. **51**(10), 1069–1077 (2013)
7. Kawakami, H., Sugano, N., Yonenobu, K., Yoshikawa, H., Ochi, T., Hattori, A., Suzuki, N.: Ef-fects of rotation on measurement of lower limb alignment for knee osteotomy. J. Orthop. Res. **22**(6), 1248–1253 (2004)
8. Sugano, N., Tsuda, K., Miki, H., Takao, M., Suzuki, N., Nakamura, N.: Dynamic meas-urements of hip movement in deep bending activities after total hip arthroplasty using a 4-dimensional motion analysis system. J. Arthroplasty **27**(8), 1562–1568 (2012)
9. Miki, H., Sugano, N., Yonenobu, K., Tsuda, K., Hattori, A., Suzuki, N.: Detecting cause of dislo-cation after total hip arthroplasty by patient-specific four-dimensional motion analysis. Clin. Biomech. **28**, 182–186 (2013)
10. Damsgaard, M., Rasmussen, J., Christensen, S.T., Surma, E., de Zee, M.: Analysis of musculoskeletal systems in the AnyBody Modeling System. Simul. Model. Pract. Theor. **14**(8), 1100–1111 (2006)
11. Bolsterlee, B., Veeger, D.H., Chadwick, E.K.: Clinical applications of musculoskeletal modelling for the shoulder and upper limb. Med. Biol. Eng. Comput. **51**(9), 953–963 (2013)

12. Arnold, E.M., Ward, S.R., Lieber, R.L., Delp, S.L.: A model of the lower limb for analysis of human movement. Ann. Biomed. Eng. **38**(2), 269–279 (2010)
13. Christophy, M., Senan, N.A.F., Lotz, J.C., O'Reilly, O.M.: A musculoskeletal model for the lumbar spine. Biomechan. Model. Mechanobiology **11**(1–2), 19–34 (2012)

Augmented Reality for Specific Neurovascular Surgical Tasks

Marta Kersten-Oertel[1,3(✉)], Ian J. Gerard[1,3], Simon Drouin[1,3], Kelvin Mok[3], Denis Sirhan[2,3], David S. Sinclair[2,3], and D. Louis Collins[1,2,3]

[1] Biomedical Engineering, McGill University, Montréal, Canada
marta.kersten@mail.mcgill.ca
[2] Neurology and Neurosurgery, McGill University, Montréal, Canada
[3] Montreal Neurological Institute and Hospital, Montréal, Canada

Abstract. Augmented reality has the potential to aid surgeons with particular surgical tasks in image-guided surgery. In augmented reality (AR) visualization for neurosurgery, the live view of the surgical scene is merged with preoperative patient data, aiding the surgeon in mapping patient images from the image-guidance system to the real patient. Furthermore, augmented reality visualization allows the surgeon to see beyond the visible surface of the head or brain at the anatomy that is relevant at different stages of surgery. In this paper, the particular surgical tasks that have benefited from AR visualization by the neurosurgeons that have used our system are described. These tasks include: tailoring a craniotomy, localizing the anatomy of interest, planning a resection corridor and determining a surgical strategy. We present each of these surgical tasks and provide examples of how AR was used in the operating room.

Keywords: Augmented reality · Neurosurgery · Neurovascular surgery · Task-based analysis · Image-guided neurosurgery

1 Introduction

Augmented reality (AR) visualization is increasingly being studied for use in the clinical domain. In AR, virtual elements and data (e.g. patient imaging data) are merged with the real world (e.g. the actual patient) in order to present information that is not readily visible or accessible to a clinical practitioner. Many research groups are exploring the use AR visualization for image-guided surgery as it promises advantages over traditional image guided surgery techniques. One of the main advantages of AR is that the patient images and surgical plans from the navigation system are visualized in the context of the real patient anatomy. In other words, in using AR the surgeon has all of the preoperative and/or intraoperative patient data (i.e. virtual objects) merged with the surgical field in a single view.

One of the main elements that has been missing from this field of research is validation of AR image-guided surgery (IGS) systems in real clinical cases in the operating room (OR) [1]. The constraints of the OR (sterility, ethics, etc.), accessibility to surgeons and operating rooms, and challenge of determining metrics for validating

C.A. Linte et al. (Eds.): AE-CAI 2015, LNCS 9365, pp. 92–103, 2015.
DOI: 10.1007/978-3-319-24601-7_10

visualization techniques make it difficult to evaluate AR visualization in image-guided surgery systems. Over the last couple of years we have brought our AR image-guided neurosurgery system (IGNS) [2, 3] into the OR to determine the feasibility of this type of visualization in neurovascular surgery. Prior to conducting an extensive clinical evaluation and validation of augmented reality image-guidance, it is important to determine the different surgical stages and tasks where augmented reality can benefit the surgeon, the operating room staff, and/or the patient. Based on the experiences of the surgeons that have used our AR navigation system in neurovascular surgery, in this paper we define the surgical tasks for which this type of visualization is useful. In order to create effective and understandable AR visualizations, it is important to focus on how to visualize the data based on the needs of the surgeon in the OR for a particular surgical task. For this reason, as well as discussing the surgical tasks that benefit from AR visualization we also describe the planning and visualization tasks that are done preoperatively in order to create useful AR views. Furthermore, we present examples of AR use from real surgical cases for each of the defined tasks.

2 Augmented Reality for Neurovascular Surgery

In the following section the types of neurovascular surgeries for which our system was used, the system itself and related work are briefly described.

2.1 Neurovascular Surgery

The focus of our work has been on developing AR visualization techniques for neurovascular surgery. Neurovascular surgery aims at managing and treating cerebrovascular disease such as aneurysms and vascular malformations. Our augmented reality system has been used during surgeries for (1) clipping aneurysms (balloon-like bulges of the vessels) and (2) resecting arteriovenous malformations (AVMs) and arteriovenous fistulae (AVF). AVMs and AVFs are abnormal tangles of blood vessels that are fed by one or more arteries (feeders) and drained by one or more veins (drainers). In the case of an AVM these tangles occur in the brain or spine, whereas AVFs occur in the dura mater.

2.2 AR IGNS System

The developed augmented reality system consists of three components: a neuronavigation workstation, a camera (either a Sony HDR XR150 video camera or a Point Grey Firefly MV USB camera that is attached to an ocular of the neurosurgical microscope), and a Polaris tracking system (Northern Digital, Waterloo, Canada) that tracks the patient, surgical tools and the camera. The neuronavigation workstation runs Ubuntu 12.04 (64-bit), with an Intel Core i7-3820@3.6 GHz on a quad-core processor with 32 GB RAM. The graphics card is a GeForce GTX 670 and the video capture card is a Conexant cx23800. The custom-built neuronavigation software, IBIS, is written in C++ and uses the Visualization Toolkit (version 5.10), the Qt 4 user interface platform, and the Insight Registration and Segmentation Toolkit (version 4.4).

To create the AR view, live images of the surgical scene are captured by the calibrated camera and augmented on the IBIS navigation system. In order to merge the preoperative patient data with the live view of the surgical scene three things are necessary: the camera must be calibrated, it must be tracked it in the OR and a patient-to-image registration must be done. For calibration, intrinsic (e.g. focus and image center) and extrinsic (i.e. the transform between the camera tracker and the optical center of the camera) parameters are computed by taking images of a checkerboard pattern and using an implementation of Zhang's method [4] to determine the transform between image and real world coordinates. Patient-to-image registration, which determines the transformation between the patient and the preoperative images is done in the operating room using corresponding anatomical landmarks (e.g. the bridge of the nose) as described in [5]. The camera calibration matrix and patient-to-image registration gives us the mapping between the preoperative images and the live images of the surgical field of view, allowing for the creation of the AR visualization.

At any point in time during the surgery, the surgeon can make use of AR visualization for surgical decision-making and guidance. In Fig. 1 we show the system being used in the operating room. For a detailed description of the system, calibration and registration procedure, we refer the reader to [3].

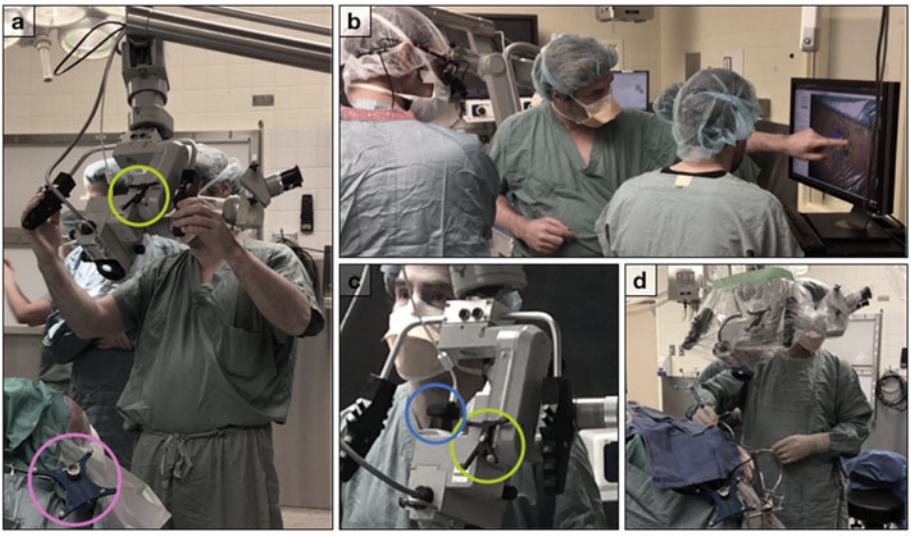


Fig. 1. The IBIS augmented reality system in use in the operating room. A tracker is attached to the neurosurgical microscope (indicated in green circle in (a) & (c)) and a Point Grey Firefly MV USB camera is mounted onto one of the oculars of the microscope (indicated in blue circle in (c)), it is calibrated and used to capture the live views of the surgical scene. The tracker gives the position of the camera with respect to the cranial reference (indicated in pink circle in (a)). In (b) the surgeon looks at the AR visualization and discusses the anatomy with the resident in order to plan the surgical approach prior to incision. During surgery, the surgeon can make use of the AR image-guidance at any time (d) (Color figure online).

2.3 Related Work

Neurosurgery was one of the first applications of clinical AR and is currently one of the most popular [1]. In the early 90 s, Gleason *et al*. [6] proposed the first neurosurgical AR system, which combined 3D segmented virtual objects (e.g. tumours) from preoperative patient images with live video images of the patient. Edwards *et al.* [7, 8] developed a microscope-assisted guided intervention (MAGI) neuronavigation system that allowed for stereo projection of virtual images into a neurosurgical microscope for ear, nose and throat (ENT) surgery and neurosurgery. Birkfellner *et al.* [9, 10] developed the Varioscope AR, a custom-built head-mounted operating microscope that allowed for virtual objects to be presented to the viewer using VGA displays. In the particular domain of neurovascular surgery, Cabrilo *et al*. used the Zeiss OPMI Pentero's Multivision function that injects virtual images into one ocular of the neurosurgical microscope to carry out two studies, one for AVMs [11] and one for aneurysms [12]. Based on feedback from the surgeons, their results suggested that AR visualization enabled a more tailored surgical approach and optimal clipping of aneurysms. In AVM surgery, the authors found that AR was useful for tailoring craniotomies, guiding dissections, and localizing draining arteries; however, it did not provide useful information in terms of feeding arteries of the AVM. For more information as to the use of augmented reality in image-guided surgery the reader is referred to [1].

3 Methods

In the following section we describe the methods for preparing and using the AR IGNS system in the operating room. In particular we look at those sets of tasks that are done prior to surgery and those that are done during surgery as shown in Fig. 2.

3.1 Preoperative Tasks: Image Processing, Visualization and Planning

In order to prepare for each case, the preoperative imaging data is processed to identify all structures and regions of interest. The data is visualized for the surgeon during preoperative planning and may be updated based on the surgeons needs for a given surgical case. A scene with the data and any planning information (e.g. landmarks for registration and selected anatomical landmarks/vessels of interest) is saved for use in the operating room.

Image and Visualization Processing. One of the main goals of AR visualization is to provide information that is not readily available or that is not obvious in the live surgical scene. In neurovascular surgery, localization of key vessels (e.g. deep feeding arteries) and intraoperative differentiation between arteries and veins in malformations can be challenging. Therefore, preoperatively we focus on identifying vessels of interest that relate to the malformation using landmarks and colour-coding based on vessel type. By developing visualization that supports anatomical understanding, we can better aid clinical decision-making and improve image-guidance.

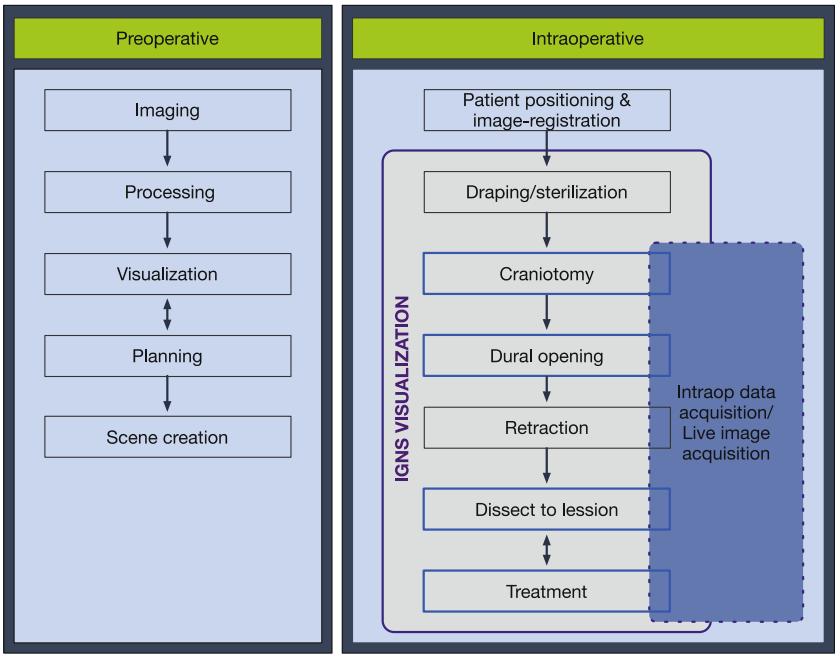


Fig. 2. The main tasks that are in association with preoperative planning and intraoperative IGNS use as well as the main surgical tasks. Tasks where we have found AR useful are indicated with blue (Color figure online).

The data that is most often acquired and used for planning and guidance in neurovascular surgery at the Montreal Neurological Institute & Hospital is computed tomography angiography (CTA). With CTA imaging, a contrast substance is injected into the patient allowing for imaging of the arteries and veins. Three phases are made available for navigation: arterial, capillary and venous.

Typically, the arterial and venous phases are colour coded and visualized for the surgeon during planning. If MRI is also available for the patient, the cortical surface is extracted [13] and rendered for planning. Based on the images, the surgeon will typically pick points of interest on the visualized vasculature, for example, feeding arteries, draining veins, and other vessels that may be pertinent during surgery (Fig. 3). A colour scheme is developed based on the available data and needs of the surgeon. Furthermore, depending on the case either volume rendering or surface rendering is used. In cases where automatic flow information causes mislabeling of vessels (arteries identified as veins or vice-versa) [3], manual segmentation and colour coding is done using MeshLab[1] (e.g. as seen is Fig. 4). We have found that preoperative vessel evaluation with the surgeon is of utmost importance in identifying all the important vessels and their direction of flow, particularly for vascular malformations.

[1] http://meshlab.sourceforge.net

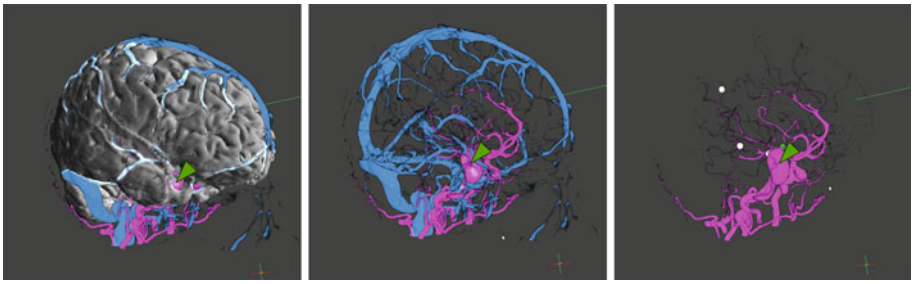


Fig. 3. Planning scene for a craniotomy for a right frontal aneurysm (indicated with green triangle) fed by the middle cerebral artery. The surgeon chose landmarks on arteries feeding the aneurysm, as well as landmarks on a superficial vessel in case an extracranial to intracranial (EC-IC) bypass was needed. Arterial phase is encoded in pink, venous in blue (Color figure online).

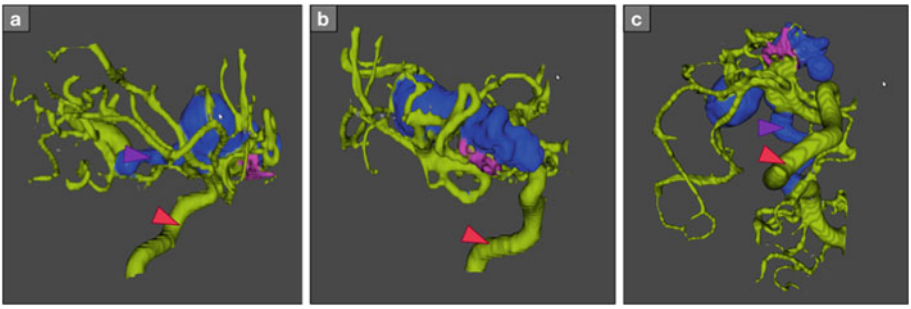


Fig. 4. Manually colour coded surface mesh for a craniotomy for an AVM (right lateral (a), anterior-posterior (AP) (b) and inferior (c) views). The red arrow indicates the carotid artery and purple arrows the draining vein. The main draining vein and associated venous aneurysm are coloured blue, and the AVM feeding arteries are coloured pink. Other vessels that are not part of the malformation are colour coded as green (Color figure online).

The plan that is developed with the surgeon (including datasets to be used, colour coding, points of interest, and landmarks for patient-to-image registration) is then saved so that it may be loaded in the operating room.

3.2 Intraoperative Tasks: Surgical and IGNS

During surgery, the typical tasks involved in image-guided procedures are performed. As described in Fig. 2, the patient is first anaesthetized and positioned and then a patient-to-image registration is done. In our case this is a landmark registration procedure that is done simultaneously with the registration procedure required for the commercial neuronavigation system, a Medtronic StealthStation (Dublin, Leinster, Republic of Ireland). Furthermore, the Medtronic and our AR system track the same

surgical tools; this ensures that there are no additional steps in the typical workflow of setting up the AR system.

Next the surgeon plans and performs the craniotomy. After opening the dura the surgeon may retract some tissue and dissect down to the anatomy of interest to treat the malformation by resecting, clipping and obliterating vessels. Traditional image-guidance visualization or AR guidance (which requires capturing live images of the surgical scene) can be used during any of these surgical tasks to aid surgical decision-making and guidance. In some cases intraoperative imaging in the form of ultrasound is used to account for brain shift and to create a more accurate overlay between real and virtual images in the augmented reality view [14].

Data Collection. During the surgical cases each use of augmented reality and all comments from the surgeon are recorded. The use of the commercial neuronavigation system and whether the PACS system was accessed for more data is also noted. The surgeon is asked to comment on each case and the use of AR for different tasks, in terms of visualization and in terms of benefit. For the first 8 cases, the surgeon was asked to use augmented reality whenever he thought it might be beneficial. Based on those cases we determined the main tasks for which AR was useful and describe these below. On the most recent cases and in all future cases the surgeon will fill out a questionnaire rating the usefulness of AR for the identified surgical tasks on a Likert scale.

4 Results and Discussion

In the following section we report the different tasks that the surgeons who have used our system have determined as benefiting from augmented reality visualization (Table 1). We describe each of the tasks and present examples from surgical cases of how AR was used in the operating room.

Table 1. Surgical tasks for neurovascular surgery and the possible uses of AR.

Surgical Tasks	Uses of AR
Craniotomy	Planning skin incision, planning size and shape of skull to remove (craniotomy), planning dural incision, identifying/tracing superficial and deep vessels
Dissection	Identifying feeding arteries, draining veins, and aneurysms, planning corridor to anatomy of interest
Vessel Identification	Differentiate between artery and vein
Treatment	Assessing normal vasculature anatomy post resection of AVM/AVF, Aiding in clip choice for aneurysm obliteration, Clipping/obliterating abnormal vessels

4.1 Tailoring the Craniotomy

The surgeons have consistently found that augmented reality is very useful for planning the craniotomy (i.e. the removal of the skull bone to expose the brain). Planning a craniotomy involves determining: (1) the size and site of the skin incision, (2) the actual size and shape of the bone to be removed and (3) opening the dura. The AR system is used at each of these points – on the skin, on the bone and on the dura. At this stage during surgery, the surgeon uses augmented reality to view the extent of a vascular malformation and its location relative to his view of the patient. The visualization of the anatomy of interest below the visible surface of the skin or skull allows the surgeon to determine the most appropriately-sized craniotomy for each surgical case. In Fig. 5, augmented reality images of the skin, bone and dura, recorded during a craniotomy for a large aneurysm case are shown. The surgeon used AR to determine the location of the aneurysm and the feeding arteries and their position with respect to the skin and bone surface to plan the craniotomy. In this particular surgical case, a surface artery was identified, skeletonized and reflected posteriorly in case an extracranial to intracranial (EC-IC) bypass was needed. To better plan the craniotomy in order to facilitate access to this vessel, AR was used to trace the superficial temporal artery branch.

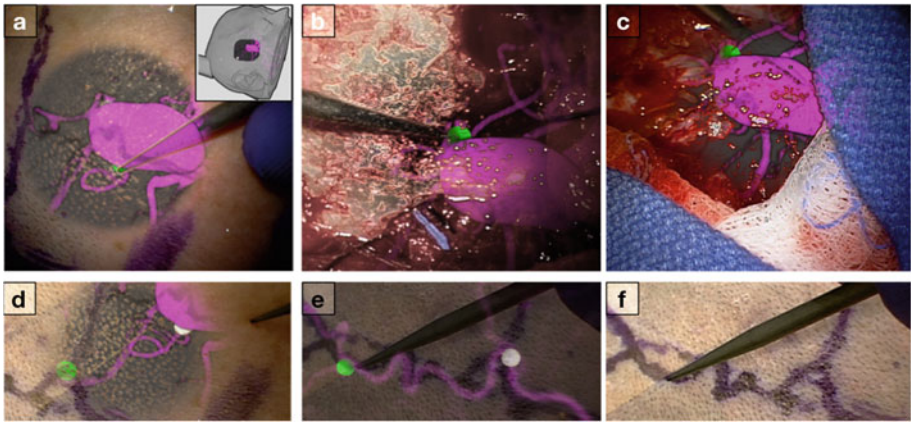


Fig. 5. Augmented reality for tailoring the craniotomy is used on the skin, bone and dura. The surgeon determined the location of important arteries (arising from the aneurysm) indicated as a green dot (a, b, c). AR was also used to determine the location of the superficial temporal artery (STA) that was harvested in case a bypass to this artery was needed (d). The course of the STA was traced on the skin using AR visualization (e, f) (Color figure online).

We have also found that prior to the craniotomy, the surgeon finds it beneficial to manipulate the augmented reality imaging to explain and show the different aspects of the vascular malformation or vessels to residents or students that may be present in the OR. Showing the actual anatomical and vascular areas of interest projected onto the actual patient who is asleep and positioned on the operating room table facilitates

discussion of the surgical plan. We plan to explore the spectrum of AR for teaching in future work.

4.2 Planning Dissection

In order to plan a dissection corridor, a surgeon can use AR visualization to determine the location of the anatomy of interest and determine eloquent areas that should be avoided. AR allows the surgeon to see the anatomy of interest that is not visible on the exposed surface of the patient and determine the best approach. In Fig. 6 we show the surgeon placing Surgicel (Ethicon, Johnson & Johnson, US), a hemostatic agent on top of a virtual marker of a deep feeding artery to an AVM. Based on this localization the surgeon was able to plan an optimal dissection corridor to access the feeding artery.

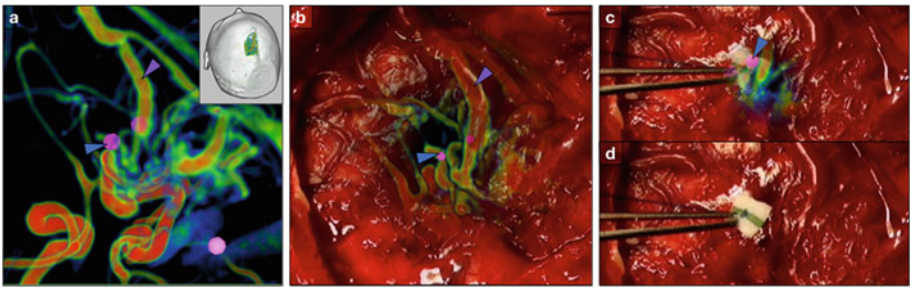


Fig. 6. Augmented reality visualization of a craniotomy for an AVM. Feeding vessels to the AVM were marked with pink points during planning (a). In order to plan a corridor of dissection the surgeon placed Surgicel, a hemostatic agent, on the cortex (b) above the location of a deep feeding artery that was to be clipped first (c, d), indicated with blue triangle in (c). In (a) and (b) we see a large draining vein (indicated in purple) that was mislabeled as an artery using 4D CTA data (Color figure online).

4.3 Vessel Identification

In neurovascular surgery, malformations can make intraoperative identification of a vessel as artery or vein difficult and determining the direction of blood flow is of utmost importance. For example, in arteriovenous malformations or fistulae, the surgical plan involves first clipping or obliterating all feeding vessels to the nidus, then clipping draining veins, and lastly safely removing the nidus. Identification and differentiation between feeding and draining vessels is often difficult intraoperatively because these vessels are neither true arteries nor true veins. Augmented reality visualization, where colour coding can be used to differentiate between arteries and veins, facilitates this task. We have found that by colour coding and using landmarks we can enable the surgeon to better differentiate, intra-operatively, those veins that mimic arteries and *vice versa*. In these special 'fistula' cases, automatic flow techniques that rely on 4D CTA do not properly label vessels because of high velocity blood flow and shunting of blood through the malformations that cause veins to appear in arterial phases.

Therefore, manual segmentation based on preoperative study and individual vessel identification that relies upon the surgical and anatomical knowledge of the clinicians is necessary. In Fig. 6 we show a mislabeled draining vein of an AVM (indicated with purple arrow) where 4D CTA data was used.

In contrast to the work of Cabrilo *et al*. [11], the surgeons that have used our system have found that AR visualization aids with the localization of (deep) feeding arteries in AVMs and AVFs. This difference may be partially attributed to using more sophisticated rendering techniques as well as preoperatively choosing and placing landmarks on the feeding arteries, and using these landmarks to localize the selected vessels intra-operatively.

4.4 Planning and Executing Treatment

Planning treatment intra-operatively involves such tasks as determining the risk of treatment, deciding which vessel to expose and obliterate or clip at a given time and how to do so, determining the type and size of the clip to use, etc. These tasks are typically done after some dissection and once a pertinent vessel is exposed. We have found that after retraction of the brain and significant dissection, the system is no longer accurate and the AR view does not correspond well with the real anatomy due to large amounts of brain shift (Fig. 7). For this reason, we have not thoroughly explored the use of AR for treatment tasks. New solutions that account for deformations and brain shift are needed in order for AR visualization to be useful at later stages during surgery. We are currently exploring the use of intraoperative ultrasound to account for brain shift and create a more accurate AR overlay in neurosurgery for brain tumours [14]. We plan to explore this avenue or research in the future for neurovascular surgery.

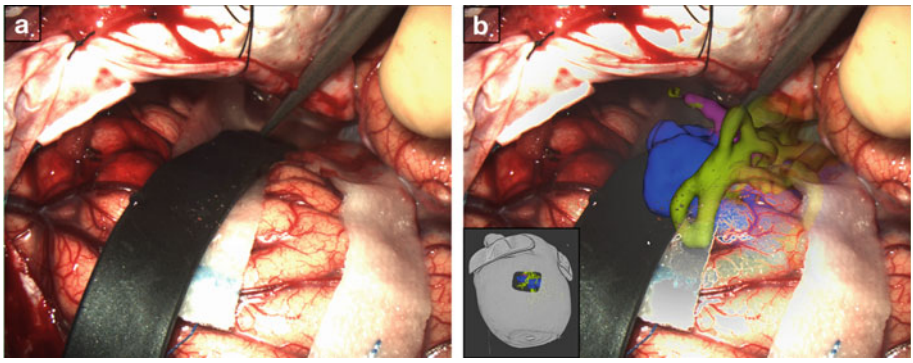


Fig. 7. Brain retraction and lesion dissection results in large amounts of surgical site deformation (a) making the AR overlay no longer accurate. Figure (b) shows frontal lobe retraction with the intracerebral/intraparenchymal AVM vessels, augmented in the view, rising into free air.

In the work of Cabrilo *et al.* [12], the authors found that AR visualization, which showed the extent and size of an aneurysm aided surgeons in determining an

appropriate hemoclip to use. To date, we have not explored the use of AR for this particular task but plan to explore it in future work.

5 Conclusions

We have used a task-based approach to determine the usefulness of augmented reality in neurovascular surgery based on experiences from the operating room and our discussions with neurovascular surgeons that have used our AR system. By determining the tasks where AR is most applicable in neurovascular surgery we can better determine validation and evaluation criteria for testing AR in each of these tasks. For example, one could imagine determining if there is a difference between traditional image-guidance and AR guidance for the task of craniotomy planning. The surgeon would first plan the craniotomy using the commercial IGNS system and next the surgeon would use the AR view. This comparison would help to determine if the surgeon changes the planned craniotomy based on the AR view. The next step of this work will be to perform rigorous studies that quantitatively determine the effect of AR for each of the described surgical tasks both on the surgeon and the patient.

References

1. Kersten-Oertel, M., Jannin, P., Collins, D.L.: The state of the art of visualization in mixed reality image guided surgery. Comput. Med. Imaging Graph. **37**, 98–112 (2013)
2. Kersten-Oertel, M., Gerard, I., Drouin, S., Mok, K., Sirhan, D., Sinclair, D., Collins, D.L.: Augmented reality in neurovascular surgery: first experiences. In: Linte, C.A., Yaniv, Z., Fallavollita, P., Abolmaesumi, P., Holmes III, D.R. (eds.) AE-CAI 2014. LNCS, vol. 8678, pp. 80–89. Springer, Heidelberg (2014)
3. Kersten-Oertel, M., Gerard, I., Drouin, S., Mok, K., Sirhan, D., Sinclair, D.S., Collins, D.L.: Augmented reality in neurovascular surgery: feasibility and first uses in the operating room. Int. J. Comput. Assist. Radiol. Surg., 26 Feb 2015
4. Zhang, Z.: A flexible new technique for camera calibration. IEEE Trans. Pattern Anal. Mach. Intell. **22**, 1330–1334 (2000)
5. Gerard, I.J., Hall, J.A, Mok, K., Collins, D.L.: New protocol for skin landmark registration in image-guided neurosurgery: technical note. Neurosurgery (2015)
6. Gleason, P.L., Kikinis, R., Altobelli, D., Wells, W., Alexander 3rd, E., Black, P.M., Jolesz, F.: Video registration virtual reality for nonlinkage stereotactic surgery. Stereotact. Funct. Neurosurg. **63**, 139–143 (1994)
7. Edwards, P., Hawkes, D., Hill, D., Jewell, D., Spink, R., Strong, A., Gleeson, M.: Augmentation of reality using an operating microscope for otolaryngology and neurosurgical guidance. J. Image Guid. Surg. **1**, 172–178 (1995)
8. Edwards, P.J., King, A.P., Hawkes, D.J., Fleig, O., Maurer Jr., C.R., Hill, D.L., Fenlon, M. R., de Cunha, D.A., Gaston, R.P., Chandra, S., Mannss, J., Strong, A.J., Gleeson, M.J., Cox, T.C.: Stereo augmented reality in the surgical microscope. Stud. Health Technol. Inform. **62**, 102–108 (1999)

9. Birkfellner, W., Figl, M., Matula, C., Hummel, J., Hanel, R., Imhof, H., Wanschitz, F., Wagner, A., Watzinger, F., Bergmann, H.: Computer-enhanced stereoscopic vision in a head-mounted operating binocular. Phys. Med. Biol. **48**, N49–N57 (2003)
10. Birkfellner, W., Figl, M., Huber, K., Watzinger, F., Wanschitz, F., Hummel, J., Hanel, R., Greimel, W., Homolka, P., Ewers, R., Bergmann, H.: A head-mounted operating binocular for augmented reality visualization in medicine - design and initial evaluation. IEEE Trans. Med. Imaging **21**, 991–997 (2002)
11. Cabrilo, I., Bijlenga, P., Schaller, K.: Augmented reality in the surgery of cerebral arteriovenous malformations: technique assessment and considerations. Acta Neurochir. (Wien) **156**, 1769–1774 (2014)
12. Cabrilo, I., Bijlenga, P., Schaller, K.: Augmented reality in the surgery of cerebral aneurysms: a technical report. Neurosurgery **10**(Suppl. 2), 252–260 (2014). discussion 260–261
13. Eskildsen, S.F., Østergaard, L.R.: Active surface approach for extraction of the human cerebral cortex from MRI. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4191, pp. 823–830. Springer, Heidelberg (2006)
14. Gerard, I.J., Kersten-Oertel, M., Petrecca, K., Drouin, S., Mok, K., Nigris, D.D., Arbel, T., Collins, D.L.: Improving augmented reality visualization with intra-operative ultrasound in image guided neurosurgery: case report. In: Cars 2015, Barcelona (2015)

Layer Separation for Vessel Enhancement in Interventional X-ray Angiograms Using Morphological Filtering and Robust PCA

Hua Ma[1(✉)], Gerardo Dibildox[1], Jyotirmoy Banerjee[1], Wiro Niessen[1,2], Carl Schultz[3], Evelyn Regar[4], and Theo van Walsum[1]

[1] Biomedical Imaging Group Rotterdam, Departments of Radiology and Medical Informatics, Erasmus MC, Rotterdam, The Netherlands
h.ma@erasmusmc.nl
[2] Quantitative Imaging Group, Faculty of Applied Sciences, Delft University of Technology, Delft, The Netherlands
[3] Department of Cardiology, Royal Perth Hospital, Perth, Australia
[4] Department of Cardiology, Erasmus MC, Rotterdam, The Netherlands

Abstract. Automatic vessel extraction from X-ray angiograms (XA) for percutaneous coronary interventions is often hampered by low contrast and presence of background structures, e.g. diaphragm, guiding catheters, stitches. In this paper, we present a novel layer separation technique for vessel enhancement in XA to address this problem. The method uses morphological filtering and Robust PCA to separate interventional XA images into three layers, i.e. a large-scale breathing structure layer, a quasi-static background layer and a layer containing the vessel structures that could potentially improve the quality of vessel extraction from XA. The method is evaluated on several clinical XA sequences. The result shows that the proposed method significantly increases the visibility of vessels in XA and outperforms other background-removal methods.

1 Introduction

Percutaneous coronary intervention (PCI) is a minimally invasive procedure for treating patients with advanced coronary artery disease. It is usually performed under guidance of X-ray angiograms (XA) where coronary arteries are opacified with contrast agent. Automatic processing of XA images, e.g. vessel extraction of coronary arteries, may serve as a basis for further processing, such as coronary motion analysis [1] and pre/intra-operative information fusion [2].

Hessian-based vessel enhancement filtering, e.g. Frangi vesselness filter [3], is commonly used for extraction of vessels in medical images. Applying such filters directly on interventional XA, however, often also enhances non-vascular structures, such as catheter segments and vertebral contours, due to their tubular or curvilinear structural appearances.

Related works have reported on methods to remove non-vessel structures or improve the visibility of vessels in XA images. In [4], a method that subtracts the

C.A. Linte et al. (Eds.): AE-CAI 2015, LNCS 9365, pp. 104–113, 2015.
DOI: 10.1007/978-3-319-24601-7_11

median frame was used for removing static structures in XA, such as vertebral bodies. Schneider et al. [5] proposed a post-processing technique on vesselness images that combines a local probability map with local directional vessel information for artifact reduction and catheter removal. Layer separation methods provide an alternative way of vessel enhancement. In [6], a multi-scale framework was developed to separate XA images into three layers based on different motion patterns such that coronary arteries are better visible in the fast motion layer. This method involves human-interactions to label corresponding control points in XA images for motion field estimation. In another study [7], a Bayesian framework was developed that combines dense motion estimation, uncertainty propagation and statistical fusion to achieve motion layer separation. Both layer separation methods require to compute motion field. Robust principal component analysis (Robust PCA) is a data decomposition technique that has e.g. been used for background modeling from surveillance video in [8]. In [9], Robust PCA was adopted for registration of DCE MR time series.

In this paper, we propose an automatic method to robustly separate foreground (contrast-enhanced vessels, guiding catheter tip) from (quasi) static background, such as vertebral bodies and guiding catheters in the aorta, while ignoring large-scale motion such as diaphragm movement. Our contributions are three folds: (1) the development of a Robust PCA based layer separation method that does not require computation of the motion field; (2) qualitative and quantitative evaluations on four clinical XA sequences; (3) comparison to other related background-removal approaches.

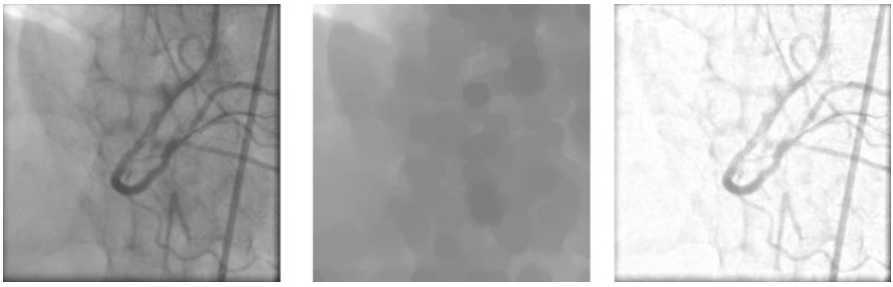
2 Method

The method enhances vessels in XA images by separating an image into three layers, i.e. a large-scale breathing layer, a quasi-static background layer and a foreground layer containing the vessels. To this end, our proposed method consists of two steps: first, separation and removal of large-scale breathing structures, such as diaphragm, from the original images, using morphological closing; second, separation of a quasi-static background from the moving structures using Robust PCA. In the remainder of this section, we describe both steps in more details, followed by the integrated layer separation.

2.1 Separation of Breathing Structures

To obtain a separate layer containing large-scale structures, we remove small objects from the original image, including guiding catheters, guide wires, stitches and vertebral bodies. Similar to the approach in [10], we apply morphological closing to the image with a circular structuring element of 8.5 mm in diameter. Pilot experiments indicated that this size was adequate for a complete removal of vessels and guiding catheters from our images while not causing too much circular artifacts. An example of a resulting image is shown in Fig. 1(b). Compared to the original image, the guiding catheter and coronary arteries are removed

and vertebral contours are blurred, while structures that presents respiratory motion, such as the diaphragm and lung tissue, remain in the image (white area in the upper left part of the image). The resulting image that contains large scale structures which exhibit respiratory motion is called the breathing layer, and will later be subtracted from the original image to obtain the difference image (DI, Fig. 1(c)) of an XA frame for further processing.



(a) (b) (c)

Fig. 1. Morphological closing operation on an XA image: (a) the original image, (b) image processed with morphological closing, (c) the difference image (DI) of (a) and (b).

2.2 Background Separation Using Robust PCA

Robust PCA decomposes a data matrix into two different sources: a low-rank matrix and a sparse matrix. Suppose that M is an $m \times n$ matrix to be decomposed, which contains n observations of m dimensional data in its columns. Robust PCA is formulated as the following optimization problem [8]:

$$\begin{aligned} \text{minimize} \quad & \|L\|_* + \lambda \|S\|_1 \\ \text{subject to} \quad & L + S = M \end{aligned} \tag{1}$$

where L is a low-rank matrix and S is a sparse matrix of the same size as M. $\|L\|_*$ denotes the nuclear norm of L and $\|S\|_1$ is the L_1 norm of S. λ is the tuning parameter of regularization. Source decomposition is achieved by solving this optimization problem. In this work, we use inexact Augmented Lagrange Multiplier (ALM) method [11] to solve the problem.[1]

Robust PCA can be applied for separation of the background layer of DI from the vessel layer. The background of an XA sequence is an image series with small changes of pixel intensity containing (quasi) static structures, while the foreground, or the vessel layer, consists of moving objects. Thus, resizing the background image into a column vector and combining all these vectors from a background series together results in a low rank matrix. Likewise, the

[1] The implementation is available at http://perception.csl.illinois.edu/matrix-rank/sample_code.html.

image series of vessel layer can be modeled as a sparse matrix, as either vessels or guiding catheters take up only a small part of the whole image content. Therefore, the background layer and vessel layer of DIs can be separated by solving the Robust PCA problem.

2.3 Image Processing Pipeline of XA Layer Separation

The proposed layer separation algorithm consists of the following steps. All steps are illustrated in Fig. 2.

1. Given an XA sequence, apply morphological closing on each frame of the series, as described in Sect. 2.1. For each frame, subtract the morphological-closed image from the original image to obtain the DI.
2. Rearrange the DIs of the XA sequence to construct a matrix whose columns represent the frames. This matrix is considered as the input matrix M in Eq. 1.
3. Solve the Robust PCA problem to obtain the background layer matrix L and vessel layer matrix S. Resize L and S to get the background layer and vessel layer of the previous size for each frame of the sequence.

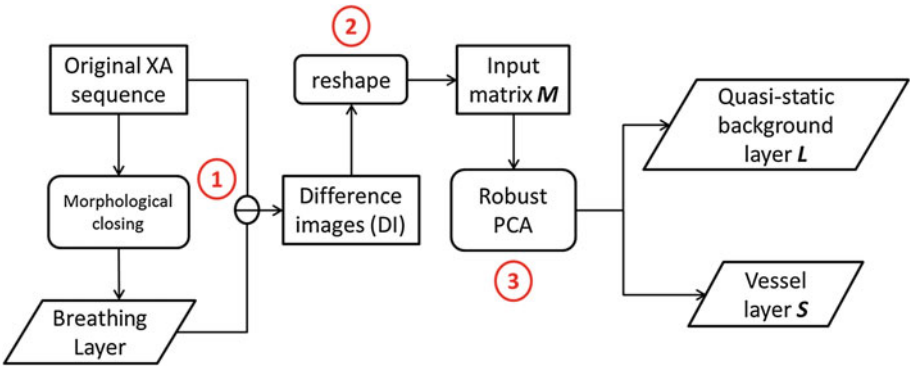


Fig. 2. The pipeline of the proposed layer separation method.

3 Experiments

Fully anonymized imaging data were used in our experiments. Four XA image series that were acquired with Siemens AXIOM-Artis biplane system were analyzed. The frame rate of all sequences is 15 frames per second. The number of frames per series ranges from 55 to 169. From our data, the image matrix is 512×512 pixels for one of the series and 600×600 for the other three, with resolution 0.216×0.216 and $0.184 \times 0.184\,\mathrm{mm}^2$, respectively.

To quantify the visibility of vessels in an image, the contrast-to-noise ratio (CNR) is used in the experiments. CNR is a measure of image quality based

on contrast. Once the background and foreground of an image is defined, the definition of CNR can be formulated as:

$$CNR = \frac{|\mu_F - \mu_B|}{\sigma_B} \tag{2}$$

where μ_F and μ_B are the mean of foreground and background pixel values respectively, and σ_B is the standard deviation of the background pixel values. This definition of CNR measures the contrast between the foreground and background pixel intensities in relation to the standard deviation of the background pixel intensities. Larger CNR values imply a better contrast.

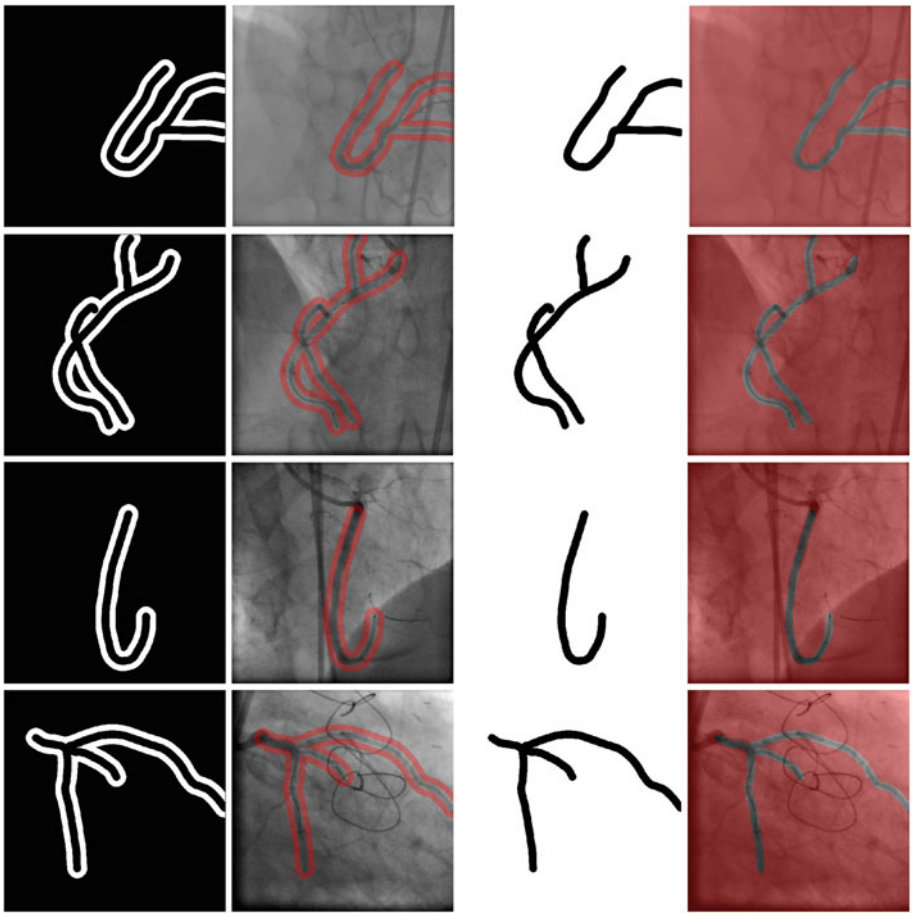


Fig. 3. Two types of mask images. Background is defined as the white image region, foreground is defined as the dark area within the white part: (Column 1) Mask 1 for one frame in the four XA sequences; (Column 2) Mask 1 overlaid on the corresponding XA frames; (Column 3) Mask 2 for one frame in the four XA sequences; (Column 4) Mask 2 overlaid on the corresponding XA frames.

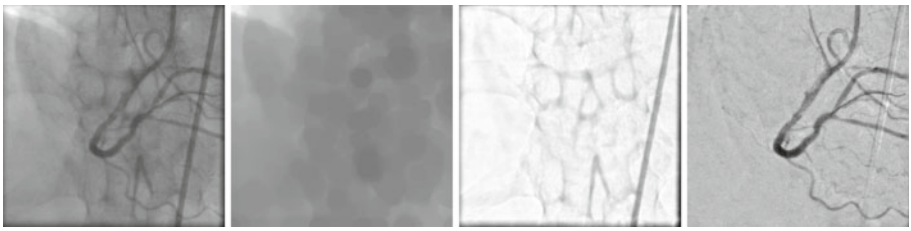
Two different versions of CNR are computed, using two different masks for defining the foreground (vessel) and the background in XA images (Fig. 3). In mask 1, as shown in Fig. 3 column 1, a 4 mm-wide image area around the manually-labeled vessel centerline is defined as the foreground (the dark area inside white region); the background are its 3 mm-wide neighborhood area (white region surrounding the vessel). This mask can be used to assess the local contrast around vessels in XA. In mask 2, as shown in Fig. 3 column 3, everything outside the foreground is considered background, which thus also evaluates the removal of the diaphragm, guiding catheters, etc. In our experiments, we randomly select 5 frames once from each sequence for the mask generation and compute the average CNR of the 5 frames.

We compare the performance of our approach to 3 other related methods. In [4], static background is eliminated by subtracting the median of the first 10 frames from each frame in the sequence. This method is referred to as *MedSubtract 1*. Second, we considered an advanced version of median subtraction by firstly removing the breathing layer using morphological closing and then subtracting the median. This is called *MedSubtract 2* in the experiments. Third, a conventional PCA technique is explored. The breathing layer is first removed to generate the difference image and the background layer is later reconstructed with the first principal component using PCA. This is referred to as *Normal PCA*.

For the parameter λ in the formulation of Robust PCA, we use the value suggested in [8]. All experiments were implemented in MATLAB 2013b on an Intel Core i7-4800MQ 2.70 GHz computer with 16 GB RAM running Windows.

4 Results

Figure 4 shows an example result of layer separation on one XA sequence. Note that in the original image (Fig. 4(a)), the presence of the diaphragm, the vertebral structures and the long guiding catheter segment makes extracting the vessels challenging. In the vessel layer image (Fig. 4(d)), those structures are removed, and the contrast between vessels and their neighborhood pixels is larger than in the original image.



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

Fig. 4. An example of layer separation: (a) the original image, (b) breathing layer, (c) quasi-static background layer, and (d) vessel layer.

Figure 5 presents the comparison of our proposed method (Row 5) to three other background-removal methods (Row 2–4) applied on four XA sequences. For each of the sequences, we selected a representative frame. It can be observed that all the four methods increase the visibility of vessels in XA with better

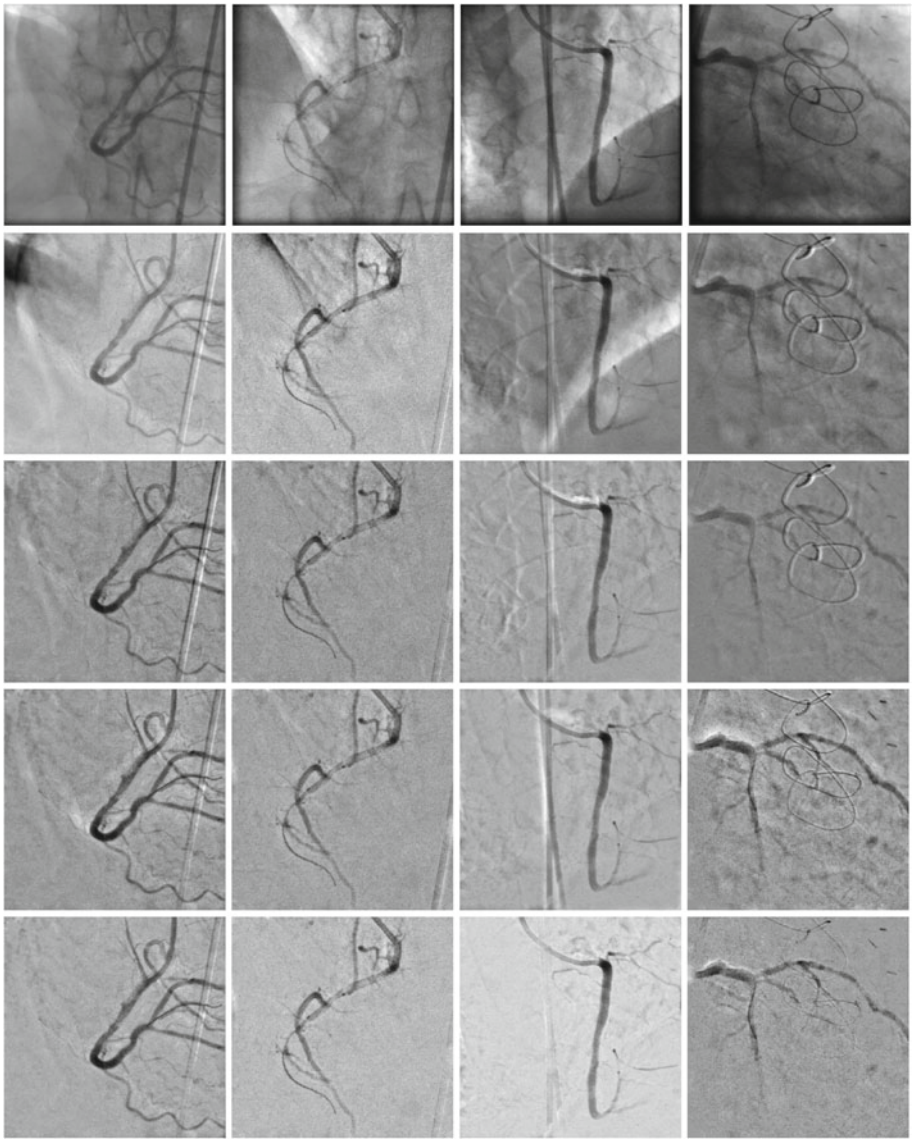
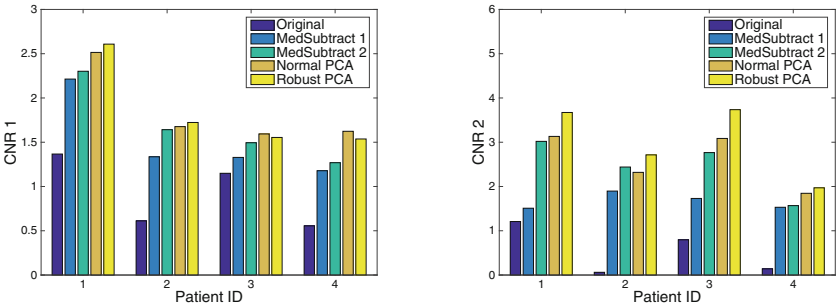


Fig. 5. Example frames of foreground images obtained by different background-removal techniques applied on four XA sequences: (Column 1–4) The four different XA sequences, (Row 1) The original image, (Row 2) *MedSubtract 1*, (Row 3) *MedSubtract 2*, (Row 4) *Normal PCA*, (Row 5) our method using Robust PCA.

contrast. However, the result of *MedSubtract 1* method (Row 2) still presents artifacts in the foreground due to the motion of diaphragm, whereas our method successfully removes the diaphragm using morphological closing. Compared to *MedSubtract 2* (Row 3) and *Normal PCA* methods (Row 4), the method based on Robust PCA performs better on removing quasi-static structures, such as the guiding catheter segment in aorta (column 1–3) and stitches (column 4).

The CNR values of XA sequences and vessel layers are illustrated in Fig. 6. Compared to the original XA, as shown in both Figs. 6(a) and (b), all methods improve the CNR values. For CNR 1, when only local contrast around vessels is measured, *Robust PCA* method performs better than the other approaches for patient 1 and 2, but has slightly lower CNR than *Normal PCA* for patient 3 and 4. In the case that the removal of diaphragm and guiding catheter is considered, as what CNR 2 indicates, *Robust PCA* is superior in all four patients.



(a) CNR 1 (b) CNR 2

Fig. 6. The average CNR over 5 randomly-chosen frames using two types of masks for the four XA sequences.

5 Discussion and Conclusion

We have developed an automatic method for layer separation of interventional XA images, to enhance vessel visualization. The method separate XA images into a breathing layer, a quasi-static background layer and a vessel layer using morphological filtering and applying Robust PCA. The separation is evaluated on four XA sequences, demonstrating better separation of the coronary arteries and reduced inclusion of breathing or quasi-static structures compared to other approaches.

Figure 5 shows that the proposed method is able to improve the visibility of vessels and performs better on representative frames of the four XA sequences. Figure 6(a) shows that the *Robust PCA* method is advantageous over the two median subtraction methods on improving the local contrast, and has similar performance with *Normal PCA*. Figure 6(b), which displays the global CNR measure, shows that *Robust PCA* is superior on all four patients which indicates

that the superiority of *Robust PCA* to other approaches is more on removing respiratory and quasi-static structures from XA to improve the contrast of vessels in the whole image. This advantage could potentially reduce the generation of spurious vessels when applying vessel extraction methods on XA.

Compared to original images, the *Robust PCA* method improves image quality in the vessel layer by removing breathing structures and background objects. Compared to the absolute-static background resulted from the median-subtraction-based methods, *Robust PCA* models a quasi-static background with small changes, which is more adaptive to the change of image content caused by coronary motion. *Normal PCA* also models a flexible background, which could be the reason why it has similar performance with *Robust PCA*. Compared to *Normal PCA*, *Robust PCA* produces less residuals of guiding catheter in the vessel layer after the removal of the background layer. The regularization parameter of *Robust PCA* enables better flexibility of balancing between moving objects and background in layer separation. Compared to other related techniques e.g. in [6,7], the main difference of the proposed method is that it does not rely on motion field, therefore, no motion field is required to extract before doing layer separation.

Several factors might have impact on CNR values. The masks defines the background and foreground, therefore the mask-related factors could directly influence the CNR values, e.g. the width of the foreground or background, whether or not including small vessels or the guiding catheter distal segment in the foreground. In addition, the number of the selected frames for mask generation from each XA sequence might also be an important factor. More in-depth analysis of these factors is part of the future work.

In conclusion, we proposed a novel layer separation method based on morphological operation and Robust PCA. We also demonstrated that the method improves the visibility of coronary arteries in XA and has advantages over several other related approaches. In the future, we will assess this technique in prospective settings and study its application in approaches that improve image guidance in XA guided cardiac interventions.

Acknowledgements. The authors wish to acknowledge financial support from Technology Foundation STW IMAGIC project under iMIT program Grant No. 12703.

References

1. Panayiotou, M., King, A.P., Housden, R.J., Ma, Y., Cooklin, M., O'Neill, M., Gill, J., Rinaldi, C.A., Rhode, K.S.: A statistical method for retrospective cardiac and respiratory motion gating of interventional cardiac x-ray images. Med. Phys. **41**(7), 071901 (2014)
2. Baka, N., Metz, C.T., Schultz, C., Neefjes, L., van Geuns, R.J., Lelieveldt, B.P.F., Niessen, W.J., van Walsum, T., de Bruijne, M.: Statistical coronary motion models for 2D+ t/3D registration of X-ray coronary angiography and CTA. Med. Image Anal. **17**(6), 698–709 (2013)
3. 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)

4. Baka, N., Metz, C.T., Schultz, C.J., van Geuns, R.J., Niessen, W.J., van Walsum, T.: Oriented gaussian mixture models for nonrigid 2D/3D coronary artery registration. IEEE Trans. Med. Imaging **33**(5), 1023–1034 (2014)
5. Schneider, M., Sundar, H.: Automatic global vessel segmentation and catheter removal using local geometry information and vector field integration. In: 2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pp. 45–48. IEEE (2010)
6. Zhang, W., Ling, H., Prummer, S., Zhou, K.S., Ostermeier, M., Comaniciu, D.: Coronary tree extraction using motion layer separation. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009, Part I. LNCS, vol. 5761, pp. 116–123. Springer, Heidelberg (2009)
7. Zhu, Y., Prummer, S., Wang, P., Chen, T., Comaniciu, D., Ostermeier, M.: Dynamic layer separation for coronary DSA and enhancement in fluoroscopic sequences. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009, Part II. LNCS, vol. 5762, pp. 877–884. Springer, Heidelberg (2009)
8. Candes, E., Li, X., Ma, Y., Wright, J.: Robust principal component analysis? J. ACM **58**(3), 11 (2011)
9. Hamy, V., Dikaios, N., Punwani, S., Melbourne, A., Latifoltojar, A., Makanyanga, J., Chouhan, M., Helbren, E., Menys, A., Taylor, S., Atkinson, D.: Respiratory motion correction in dynamic MRI using robust data decomposition registration application to DCE-MRI. Med. Image Anal. **18**(2), 301–313 (2014)
10. Ma, H., Dibildox, G., Schultz, C., Regar, E., van Walsum, T.: PCA-derived respiratory motion surrogates from X-ray angiograms for percutaneous coronary interventions. Int. J. Comput. Assist. Radiol. Surg. **10**(6), 695–705 (2015)
11. Lin, Z., Liu, R., Su, Z.: Linearized alternating direction method with adaptive penalty for low-rank representation. Advances in Neural Information Processing Systems, 612–620 (2011)

Automatic Guide-Wire Detection for Neurointerventions Using Low-Rank Sparse Matrix Decomposition and Denoising

Markus Zweng[1,3], Pascal Fallavollita[1], Stefanie Demirci[1], Markus Kowarschik[3], Nassir Navab[1,2], and Diana Mateus[1,4(✉)]

[1] Computer Aided Medical Procedures, Technische Universität München, Munich, Germany
[2] Computer Aided Medical Procedures, Johns Hopkins University, Baltimore, USA
[3] Siemens Corporate Research, Munich, Germany
[4] Institute of Computational Biology, Helmholtz Zentrum München, Munich, Germany
mateus@in.tum.de

Abstract. In neuro-interventional surgeries, physicians rely on fluoroscopic video sequences to guide tools through the vascular system to the region of interest. Due to the low signal-to-noise ratio of low-dose images and the presence of many line-like structures in the brain, the guide-wire and other tools are difficult to see. In this work we propose an effective method to detect guide-wires in fluoroscopic videos that aims at enhancing the visualization for better intervention guidance. In contrast to prior work, we do not rely on a specific modeling of the catheter (*e.g.* shape, intensity, *etc.*), nor on prior statistical learning. Instead, we base our approach on motion cues by making use of recent advances in low-rank and sparse matrix decomposition, which we then combine with denoising. An evaluation on 651 X-ray images from 5 patient shows that our guide-wire tip detection is precise and within clinical tolerance for guide-wire inter-frame motions as high as 6 mm.

1 Introduction

Neuro-interventional procedures are minimally invasive surgeries designed for the treatment of pathologies in the cerebrovascular system. Examples of neuro-interventions are endovascular aneurysm coiling, embolization of fistulae, or for stenosis, intracranial angioplasty and stenting. During such procedures surgeons insert a guide-wire through the patient's femoral artery in order to navigate the tools through the catheter up to the brain. The navigation is performed under image guidance, in particular angiography, where a sequence of fluoroscopic X-ray images show the motion of the interventional tools inside the patient at video frame-rate (typically 7.5 or 15 fps). Navigation is a complex task for several reasons: (i) the sensible anatomy in the brain, (ii) the low contrast, resolution and signal-to-noise ratio resulting from limiting the radiation dose, (iii) the fact that X-ray images are only two-dimensional projections of three-dimensional structures, causing structure super-imposition and occlusions (*cf.* Fig. 1-top left), and

C.A. Linte et al. (Eds.): AE-CAI 2015, LNCS 9365, pp. 114–123, 2015.
DOI: 10.1007/978-3-319-24601-7_12

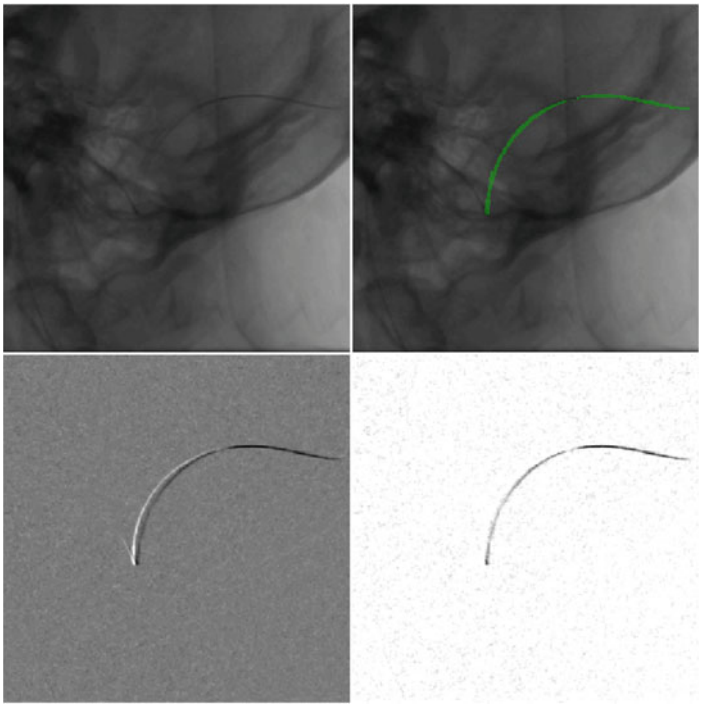


Fig. 1. Example of results for one frame, cropped and scaled for clarity. **(top left)** original frame. **(top right)** results for the proposed method (RPCA/FNLM). **(bottom left)** Estimated foreground $\mathbf{E}$. **(bottom right)** Estimated foreground $\hat{\mathbf{E}} = \mathbf{E} < 0$.

(vi) motion blur. In this work we aim at assisting the surgeon by automatically detecting and visualizing the guide-wire (see Fig. 1 top right). Such guide-wire detection could be further used in higher level visualizations, *e.g.* for 3D guidance using 3D models of the vessels acquired preoperatively with a rotational X-ray angiography [1,3], which is our long-term goal.

Prior work addressing enhanced tool visualization for navigation are based either on a frame-wise detection or on tracking [8]. We focus hereon on detection techniques which are more suitable for a fully automated solution and can serve as input or reinitialization for tracking. Most of current detection techniques exploit the prior knowledge on the line like structure of the guidewire. For instance, Petkovic *et al.* [15] designed a filter to enhance line-like structures observed through X-Ray images, while Bismuth *et al.* [4], presented a curvilinear structure enhancement using a polygonal path image. These methods are general but may have problems detecting the tortuous guide-wire tips. Furthermore, they lead only to an enhanced image possibly containing other undesired structures. To address these issues, Lessard *et al.* [12] combined filtering with segmentation, whereas Honnorat *et al.* [9] relied on steerable filters regularized by tensor-voting. However, due to the difficulty of selecting the structures

corresponding to the guidewire both methods still rely on the manual selection of segmentation seeds for the former, or the guide-wire endpoints for the later, in order to initialize the algorithm. To overcome the difficulties above, there has been an increasing interest in automatic guide-wire detection by means of learning techniques, *e.g.* [2,10,14], which have shown promising results at the cost of a prior training stage and large amounts of manually labelled data needed in order to generalize to different image settings.

In this work we propose to detect tools in fluoroscopic sequences by relying on their motion, which is a less explored type of prior knowledge. A first approach in this direction was proposed by Spiegel *et al.* [16], where a simple background subtraction is applied using as mask an initial frame where no guide-wire is present. Although such detection is fast and easy to implement, it assumes a static background; so no patient motion, illumination variations nor image-parameter changes are allowed. Inspired by advances in sparse and low-rank decomposition of matrices we propose here a two step approach to detect the guidewire during neuro-interventional procedures. In the *first step*, we use a robust background subtraction method based on Robust PCA (RPCA) [6] to detect candidate guidewire regions. Several reasons motivate our choice for RPCA: (i) robustness to variations in image brightness, (ii) ability to handle a dynamic background, and (iii) capability of accounting for the partly-correlated noise present in X-ray sequences. In the *second step*, we investigate different methods to filter remaining noise in the foreground estimation. We evaluate our solution on clinical data and show its robustness for inter-frame guidewire motion of up to 6 mm.

2 Method

The input to our method is the fluoroscopic video used for guiding a neuro-intervention. We assume the main source of visible motion in the video is caused by the interventional tools, while the remaining part of the images will be close to static and considered as background. The tools can then be separated by subtracting the background from the current image of interest. To enforce robustness against subtle motions and imaging condition changes, the background is estimated w.r.t. the previous k frames in a sliding window fashion. More formally, we denote the video frames by $\mathbf{f}_i \in \mathbb{R}^s$, $i \in \{1, \ldots, N\}$, with N the length of the video and $s = m \times n$ the size of the image. For each frame of interest $\mathbf{f}_i$, the goal is to obtain a binary mask $\mathbf{M}_i \in \{0, 1\}^s$ indicating which pixels belongs to the guidewire. Our method is composed of two steps. First, for each time-window Robust Principal Component Analysis (RPCA) [6] is applied to detect candidate guidewire regions for the current frame (*cf.* Sect. 2.1). Then, the remaining noise is filtered out (*cf.* Sect. 2.2).

2.1 Low-Rank Background Subtraction via RPCA

To estimate the background of current image $\mathbf{f}_i$, we consider a time window including the last k frames and stack them in the columns of the data matrix

$\mathbf{D} = [\mathbf{f}_{i-k}, \ldots, \mathbf{f}_i]$, where $\mathbf{D} \in \mathbb{R}^{s \times k}$. Given the neuro-interventional sequences have video frame-rates, it is expected that the contiguous frames composing $\mathbf{D}$ are highly correlated. If the correlation is linear, then the matrix $\mathbf{D}$ is low-rank and it is possible to estimate the background masks by finding a rank-q matrix $\mathbf{L}$ approximating the data matrix $\mathbf{D}$, *i.e.* by optimizing $\min_{\mathbf{L}} \parallel \mathbf{D} - \mathbf{L} \parallel$ s.t. $\text{rank}(\mathbf{L}) \leq q$, where $q \leq \text{rank}(\mathbf{D})$. This minimization can be solved using ordinary PCA over the data matrix $\mathbf{D}$ and retaining only the most significant components to build $\mathbf{L}$.

However, moving parts primary caused by the guide-wire, would cause errors when using linear methods for estimating a low-rank representation. To handle such cases, where gross but sparse corruptions are present, Candes *et al.* [6] proposed the following Robust PCA method. Assuming the data matrix $\mathbf{D}$ can be modeled as the sum of a low rank matrix $\mathbf{L}$ and some additive sparse error $\mathbf{E}$, *i.e.* $\mathbf{D} = \mathbf{L} + \mathbf{E}$, RPCA is able to exactly recover[1] $\mathbf{L}$ by solving the following convex optimization problem:

$$\min_{\mathbf{L};\mathbf{E}} ||\mathbf{L}||_* + \gamma ||\mathbf{E}||_1 \text{ s.t. } \mathbf{D} = \mathbf{L} + \mathbf{E}, \tag{1}$$

where $||\cdot||_*$ denotes the nuclear norm (enforcing low-rank), $||\cdot||_1$ is the L1 norm modeling the sparsity, and γ is a positive weighting parameter controlling the sparsity (of the foreground). From the different solvers for Eq. 1 we use the inexact augmented Lagrangian multiplier method (iALM) [13]. After solution, the low rank component $\mathbf{L}$ gives us an estimate of the background, while the moving parts of the image and the noise will be encoded in the noise component $\mathbf{E}$. Actually, the foreground estimate for the current frame corresponds to the (matrix version of the) last column of $\mathbf{E}$, which we denote here $\mathbf{E}_k$. More particularly we are interested in its negative range, which contains the motion information of the dark structures of the interventional tool. To ease later notation, let us additionally denote:

$$\hat{\mathbf{E}}(\mathbf{x}) = \begin{cases} |\mathbf{E}_k(\mathbf{x})|, \forall \mathbf{E}_k(\mathbf{x}) < 0 \\ 0 \qquad \text{otherwise} \end{cases}$$

where $\mathbf{x} \in \Omega$ is a pixel taken from the image domain $\Omega \subset \mathbb{R}^2$. Finally, $\hat{\mathbf{E}}$ is normalized to lie in the range $[0, 1]$, such that intensities belonging to the guide-wire are close to 1. After this first step, the current foreground estimate $\hat{\mathbf{E}}$ will still contain noise that we target to remove in the second stage of our method described next.

2.2 Denoising

Since our low-rank background estimation can only cover up noise that is correlated between frames we need a further step to remove remaining artifacts. To do so, we compare the following four different methods.

[1] As long as the error matrix $\mathbf{E}$ is sufficiently sparse w.r.t. the rank of $\mathbf{L}$.

Thresholding and Cluster Removal (TCR). This method, chosen for its simplicity and computationally efficiency, consists of a histogram-based thresholding heuristic followed by the removal of small connected components. First, the discrete histogram of intensities for the current frame is estimated. We assume that the catheter information is in the tail of the histogram (as most pixels belonging to background have values closer to 1 after the normalization and dominate the distribution). Therefore a threshold t_{TCR} was heuristically determined by finding the position h_{tail} of the first bin exceeding a minimum of counts and defining $t_{\mathrm{TCR}} = h_{\mathrm{tail}} + \frac{1}{3}(1 - h_{\mathrm{tail}})$. Only the values below t_{TCR} are retained. Finally, each pixel $\mathbf{x}$ is assigned to a connected component $\omega_{\mathbf{x}}$ and components with fewer than c_{TCR} elements are removed. For a pixel $\mathbf{x} \in \Omega$ this results in:

$$\mathbf{M}_{\mathrm{TCR}}(\mathbf{x}) = \begin{cases} 1 & \forall \left(\hat{\mathbf{E}}(\mathbf{x}) \leq t_{\mathrm{TCR}}\right) \wedge (\#\omega_{\mathbf{x}} \geq c_{\mathrm{TCR}}) \\ 0 & \text{otherwise.} \end{cases}$$

Hessian-Based Filtering (HES). Hessian-based methods are popular for enhancing lines for guide-wire tracking and detection [4]. We used the approach presented in [16], computing the eigenvalues ($|\lambda_1| \leq |\lambda_2|$) of the image Hessian at each pixel, and determining that a line like structure is present when the following constraints are true: $|\lambda_1| \approx 0$, $|\lambda_1| \leq |\lambda_2|$, and $|\lambda_2| > t_{\mathrm{HES}}$, where t_{HES} denotes a user-specified threshold. The Hessian is computed on the basis of image second derivatives over smoothed images, where a smoothness parameter σ_{HES} determines the scale of the used Gaussian kernel. A search is done over a range of scales, $\mathbf{r}_{\mathrm{HES}} = [r_{\min}, r_g]$, where $r_{\min}$ is a small value and r_g is an estimate of the guide-wire radius. At the end, the guidewire mask for a pixel is computed for the optimal scale $\mathbf{r}^*_{\mathrm{HES}}$ leading to the highest vesselness value λ^*_2. In sum,

$$\mathbf{M}_{\mathrm{HES}}(\mathbf{x}) = \begin{cases} 1 & \text{if } |\lambda^*_1| \approx 0,\ |\lambda^*_1| \leq |\lambda^*_2|, \text{ and } |\lambda^*_2| > t_{\mathrm{HES}} \\ 0 & \text{otherwise.} \end{cases}$$

Fast Non-local Means Filtering (FNLM). Non-local means (NLM) filtering methods, introduced by Buades *et al.* [5], are known to achieve good denoising while preserving textures and fine structures, even for high noise levels. Instead of depending on the guidewire width as the HES method, NLM depends on an estimate of the standard deviation of the image noise σ_{FNML}. Given the noisy foreground estimation $\hat{\mathbf{E}}$, the denoised value for pixel $\mathbf{x}$ is computed as a weighted average over a search region $R_{\mathbf{x}} \subset \Omega$ around $\mathbf{x}$:

$$\hat{\mathbf{E}}_{\mathrm{FNLM}}(\mathbf{x}) = \sum_{\mathbf{y} \in R_{\mathbf{x}}} w(\mathbf{x}, \mathbf{y}) \hat{\mathbf{E}}(\mathbf{y}) \tag{2}$$

The weights $w(\mathbf{x}, \mathbf{y}) = -\exp \frac{\sum_{\mathbf{t} \in \{0,\ldots,p\}^2} |\mathbf{f}(\mathbf{x}+\mathbf{t}) - \mathbf{f}(\mathbf{y}+\mathbf{t})|^2}{h}$, with a h as a filtering parameter reflect the similarity of two patches of size $p \times p$ centered at $\mathbf{x}$ and $\mathbf{y}$.

This definition enforces that higher weights are given to pixels lying on patches with similar structure to the current one. In this work we used a fast variant of NLM (FNML) that allows for an efficient computation of the weights [7]. The final mask $\mathbf{M}_{\text{FNLM}}$ is computed by thresholding the denoised image with a user-defined parameter t_{FNLM}.

Markov-Random-Field (MRF). Inspired by the work in [17] we use a MRF to enforce the guide-wire continuity. A graph is built, where the vertices represent the pixels of the image $\mathcal{V} = \{\mathbf{x}_1, \ldots, \mathbf{x}_s\}$ and edges $\mathcal{E} = \{e_{\mathbf{xy}}\}$ link pairs of neighboring nodes, that is, $e_{\mathbf{xy}}$ exists if $\mathbf{y}$ lies within a square neighbourhood $\mathbf{y} \in \mathcal{N}_{\mathbf{x}}$ around $\mathbf{x}$. The binary foreground mask $\mathbf{M}_{\text{MRF}}$ for the guidewire is obtained by optimizing the following energy according to the Ising model [11]:

$$\min_{\mathbf{M}_{\text{MRF}}} \sum_{\mathbf{x} \in \mathcal{V}} u(\mathbf{M}_{\text{MRF}}(\mathbf{x})) + \sum_{e_{\mathbf{xy}} \in \mathcal{E}} \alpha |\mathbf{M}_{\text{MRF}}(\mathbf{x}) - \mathbf{M}_{\text{MRF}}(\mathbf{y})| \tag{3}$$

The first term describes the unary potential of a pixel being foreground or background. u is set to depend on the intensity value of the foreground estimation:

$$u(\mathbf{M}(k)) = \begin{cases} \beta \hat{\mathbf{E}}(k) & \text{if } \mathbf{M}(k) = 0 \\ (1 - \hat{\mathbf{E}}(k))^2 & \text{if } \mathbf{M}(k) = 1 \end{cases} \tag{4}$$

where the parameter β controls the balance between the two unary costs. The second term in Eq. 3 describes the pairwise potential imposing constraints on the mask value for neighboring pixels; it penalizes discontinuities and discouraging noise. The parameter α, controlling the strength of this neighboring dependency was chosen to be constant and Eq. 3 was minimized using graph cuts [11].

3 Experimental Validation

The validation was performed on fluoroscopic sequences of five patient datasets with a total of 651 frames of size $s = 512 \times 512$. The sequences were acquired with a frame rate of 15 fps. Two of the datasets were acquired using a biplanar X-ray system. The tip position and guide-wire centerlines were manually annotated by two experienced observers.

Implementation Details

- For the background estimation we set the time window size to $k = 20$, which corresponds to 1.333 s at the current frame-rate.
- The sparsity parameter of RPCA to $\gamma = 1/(\sqrt{\max(m, n)})$, the tolerance of the iALM solver to 10^{-4}, and its maximum iterations to 1000.
- For TCR, the number of histogram bins is set to 1000, and the threshold t_{TCR} is computed by finding the first histogram bin containing 250 elements. Also, the minimum allowed size for a connected component is $c_{\text{TCR}} = 8$.

- For HES, we set the scale range to $\mathbf{r}_{\text{HES}} = [0.3, r_g]$, with r_g given by the specifications of the guide-wire, and fix t_{HES} to the standard deviation of the foreground image $\sigma_{\hat{\mathbf{E}}}$.
- For FNLM, we set the filtering parameter as $h = 2\sigma_{\hat{\mathbf{E}}}$, the patch size to $p = 11$, and the size of the search region $R_{\mathbf{x}}$ to 43×43. h was also used as final threshold t_{FNLM}.
- Finally, for MRF, a parameter search was performed on a subset of the data. According to the F1-measure we set $0.06 \leq \alpha \leq 0.16$, and $8 \leq \beta \leq 10$.

Validation/Results. Quantitative results of our evaluation are presented in Table 1. All of the method combinations show as desired high accuracies (ACC) and very low false positive rates (FPR). This is in part due to an imbalance of the positive and negative classes as there is a significant larger amount of pixels in the background (negatives) in comparison to those depicting the catheter (positives). Furthermore, the positive predicted values (PPV), not taking into account the negative class, are also high which is associated to a good precision. The apparently low TPR values are a consequence of the videos containing relatively long periods of no guide-wire motion, where our motion-based assumption does not hold. Indeed, the performance depends on the amount of motion, as well as how much of it is captured by the current time-window. We claim that during such static frames the navigation assistance is less relevant.

On the contrary, the most important for the surgeon is being able to see the moving tip, which has been a challenge for prior techniques. Therefore, we also measure the distances of the detected tips to the ground truth, which results in values of around $d_{\text{tip}} \approx 2.5\,\text{mm}$ for at least 3 of the methods. Moreover, if we focus on the most interesting frames containing tip motion, that is, considering frames were there ground-truth tip moves up to 6 mm and removing frames for which the tip is static, then, the tip is accurately detected (e.g. $d^*_{\text{tip}} = 1.28 \pm 1.18\,\text{mm}$). To better illustrate the discussion above we also plot in Fig. 2 the influence of the underlying motion on the estimated tip distance. The crosses on the vertical axis (tip motion = 0) indicate no motion of the ground truth tip w.r.t. the

Table 1. Quantitative evaluation: True Positive Rate (TPR), False Positive Rate (FPR), Accuracy (ACC), Positive Predictive Value (PPV/Precision), avg. tip distance d_{tip}, missed tips within ROI (80×80 px), d^*_{tip} avg. tip distance for the cases where the ground truth tip moved up to 6mm between the last and the current frame (see Fig. 2). Distances are given in mm, other values in percentages.

	ACC	PPV	FPR	TPR	$d_{\text{tip}} \pm \sigma_{\text{tip}}$	Missed tips	$d^*_{\text{tip}} \pm \sigma^*_{\text{tip}}$
RPCA/TCR	99.30	70.65	0.15	16.16	2.03 ± 2.49	12.66	1.11 ± 1.20
RPCA/HES	99.38	88.11	0.05	13.33	2.54 ± 3.19	34.82	1.64 ± 1.77
RPCA/FNLM	99.37	88.71	0.06	13.78	2.24 ± 2.51	17.88	1.28 ± 1.18
RPCA/MRF	99.33	82.59	0.09	11.90	2.92 ± 2.57	28.68	1.49 ± 1.52

previous frame, leading to the RPCA not being able to detect the guidewire, and thus resulting in high tip distance estimates. On the other hand, errors in distance estimates also increase when the tip moves very fast (tip motion > 6 mm), mainly due to motion blur. In these frames tips appear washed out, and have weak intensities that prevent their detection.

Regarding the comparative performance of the different denoising techniques, we recommend FNLM given its good accuracy (ACC) and precision (PPV). Even if HES gives similar results and has less artifacts (FPR) it has twice as many missed tips than FNLM. Additionally FNLM has the advantage of being independent of the tool width. While the performance of FNLM in terms of tip distance is slightly worse than that of TCR, FNLM produces less than half of the outliers (FPR) compared to TCR, which is for the application more relevant. In case of tension stress on the guide-wire, fast motions can occur resulting in artifacts showing a second line (Fig. 1-bottom-left). Note that our method is able to differentiate such motion (see full foreground estimation Fig. 1-bottom-right) and correctly detect the guide-wire, whereas other detection methods would produce false positives or need further processing.

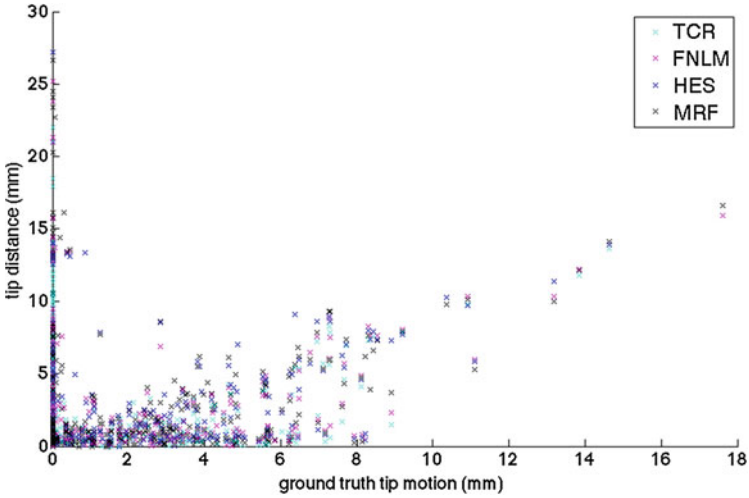


Fig. 2. Estimated tip distance vs. ground truth tip motion. The ground-tip motion is measured as the Euclidean distance of the ground-truth tip position w.r.t. the previous frame. Static tips corresponding to ground-tip motion of 0 are not well captured by our approach, as shown by the estimated distances variate along the vertical axis. This is to be expected as our method relies on motion cues. Better estimates are obtained for the four different methods, when the tool effectively moves.

4 Conclusion

Navigating the guide-wire through the vessels and positioning its tip at the operation site is a crucial step in neuro-interventional procedures. Guide-wire

tracking and detection methods have focused on the delineation of the tip or the whole guide-wire. Our approach concentrates instead on the motion information obtained from the last k frames, is independent of the interventional tool being used, and provides valuable information to the surgeon as during navigation the motion of the tip and changes in tension on vessels depict the most important pieces of information that could be overlaid.

One current limitation of the method is in detecting parts of the guide-wire that are static. This simply because our main assumption to detect the tool is that the tool moves. Although we have discussed that in a real application this assumption will be probably enough for navigation, possible ways to overcome it would be by combining our approach with an appearance model, in order to complete continuous line structures.

To conclude, we have proposed a robust guidewire detection approach that takes advantage of state-of-the-art developments in denoising and low-rank and sparse decomposition. Our work introduces a novel perspective for road-mapping in neuro-interventions that considers motion as the most important cue for tool detection. The method is precise, fully automatic, and does not require specific models of the tools, nor a training stage. It uses the image statistics over a sliding time-window of a given number of frames, thus being robust to rigid patient motion or imaging parameter changes. Finally, the algorithm requires few parameters and can be implemented to eventually achieve real-time performance. In sum, our method deals with important requirements for its real application in interventional neuroradiology.

References

1. Baert, S.A.M., van der Kraats, E.B., Niessen, W.J.: Three-dimensional guide-wire reconstruction from biplane image sequences for integrated display in 3-D vasculature. IEEE Trans. Med. Imaging **22**(10), 1252–1258 (2003)
2. Barbu, A., Athitsos, V., Georgescu, B., Böhm, S., Durlak, P., Comaniciu, D.: Hierarchical learning of curves application to guidewire localization in fluoroscopy. In: Computer Vision and Pattern Recognition (CVPR) (2007)
3. Baur, C., Milletari, F., Belagiannis, V., Navab, N., Fallavollita, P.: Automatic 3D reconstruction of electrophysiology catheters from two-view monoplane C-Arm image sequences. In: International Conference on Information Processing in Computer-Assisted Interventions (2015)
4. Bismuth, V., Vancamberg, L., Gorges, S.: A comparison of line enhancement techniques: applications to guide-wire detection and respiratory motion tracking. In: SPIE Conference Series, vol. 7529 (2009)
5. Buades, A. Coll, B., Morel, J.M.: A non-local algorithm for image denoising. In: Computer Vision and Pattern Recognition (CVPR) (2005)
6. Candès, E.J., Li, X., Ma, Y., Wright, J.: Robust principal component analysis? J. ACM **58**(3), 11:1–11:37 (2011)
7. Darbon, J., Cunha, A., Chan, T., Osher, S., Jensen, G.: Fast nonlocal filtering applied to electron cryomicroscopy. In IEEE International Conference on Biomedical imaging: from Nano to Macro, pp. 1331–1334 (2008)

8. Heibel, H., Glocker, B., Groher, M., Paragios, N., Komodakis, N., Navab, N.: Discrete tracking of parametrized curves. In: IEEE Computer Vision and Pattern Recognition (CVPR) (2009)
9. Honnorat, N., Vaillant, R., Paragios, N.: Graph-based geometric-iconic guide-wire tracking. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part I. LNCS, vol. 6891, pp. 9–16. Springer, Heidelberg (2011)
10. Honnorat, N., Vaillant, R., Paragios, N.: Robust guidewire segmentation through boosting, clustering and linear programming. In: IEEE International Conference on Biomedical Imaging: From Nano to Macro, pp. 924–927 (2010)
11. Kolmogorov, V., Zabin, R.: What energy functions can be minimized via graph cuts? IEEE Trans. Pattern Anal. Mach. Intel. (PAMI) **26**(2), 147–159 (2004)
12. Lessard, S., Lau, C., Roy, D., Soulez, G., de Guise, J.A.: Wires segmentation in fluoroscopic images during cerebral aneurysm endovascular intervention. In: IEEE International Conference on Biomedical Imaging: From Nano to Macro. IEEE (2008)
13. Lin, Z., Chen, M., Ma, Y.: The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices. Math. Program. (2010)
14. Milletari, F., Belagiannis, V., Navab, N., Fallavollita, P.: Fully automatic catheter localization in C-Arm images using $\ell 1$-sparse coding. In: Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R. (eds.) MICCAI 2014, Part II. LNCS, vol. 8674, pp. 570–577. Springer, Heidelberg (2014)
15. Petković, T., Lončarič, S.: Guidewire tracking with projected thickness estimation. In: IEEE International Conference on Biomedical Imaging: From Nano to Macro, pp. 1253–1256 (2010)
16. Spiegel, M., Pfister, M., Hahn, D., Daum, V., Hornegger, J., Struffert, T., Dörfler, A.: Towards real-time guidewire detection and tracking in the field of neuroradiology. In: SPIE Conference Series, pp. 726105–726108 (2009)
17. Zhou, X., Yang, C., Yu, W.: Moving object detection by detecting contiguous outliers in the low-rank representation. IEEE Trans. Pattern Anal. Mach. Intel. **35**(3), 597–610 (2013)

3D Surgical Overlay with Markerless Image Registration Using a Single Camera

Junchen Wang[1(✉)], Hideyuki Suenaga[2], Liangjing Yang[1], Hongen Liao[3], Takehiro Ando[1], Etsuko Kobayashi[1], and Ichiro Sakuma[1]

[1] The University of Tokyo, Tokyo, Japan
wangjunchen@gmail.com
[2] The University of Tokyo Hospital, Tokyo, Japan
[3] Tsinghua University, Beijing, China

Abstract. Minimum invasive surgery can benefit from surgical visualization, which is achieved by either virtual reality or augmented reality. We previously proposed an integrated 3D image overlay based surgical visualization solution including 3D image rendering, distortion correction, and spatial projection. For correct spatial projection of the 3D image, image registration is necessary. In this paper we present a 3D image overlay based augmented reality surgical navigation system with markerless image registration using a single camera. The innovation compared with our previous work lies in the single camera based image registration method for 3D image overlay. The 3D mesh model of patient's teeth which is created from the preoperative CT data is matched with the intraoperative image captured by a single optical camera to determine the six-degree-of-freedom pose of the model with respect to the camera. The obtained pose is used to superimpose the 3D image of critical hidden tissues on patient's body directly via a translucent mirror for surgical visualization. The image registration performs automatically within approximate 0.2 s, which enables real-time update to tackle patient's movement. Experimental results show that the registration accuracy is about 1 mm and confirm the feasibility of the 3D surgical overlay system.

1 Introduction

Three dimensional (3D) image overlay based augmented reality (AR) surgical navigation has been introduced many times in the literature [3–5,8]. The 3D image with both horizontal and vertical parallax is displayed by a lens array monitor (3D display) which can be observed without wearing glasses. In our recent work [12], we proposed an integrated 3D image overlay based surgical visualization solution including 3D image rendering, distortion correction, and spatial projection. Given an anatomical mesh model derived from CT data, the 3D image of the model, which has the same geometric dimensions as the original organ, can be rendered in real time with the help of a graphics processing unit (GPU). To superimpose the 3D image on patient's body with correct overlay, the pose of the anatomical model (i.e., the digitalized patient organ in the image

C.A. Linte et al. (Eds.): AE-CAI 2015, LNCS 9365, pp. 124–133, 2015.
DOI: 10.1007/978-3-319-24601-7_13

space) with respect to the 3D display has to be determined intraoperatively. 3D measurement systems (e.g., Polaris Tracking System) are usually employed for this purpose. The pose determination is composed of two steps. One is to calculate the transformation from the 3D display to the 3D measurement system, which is called display-camera calibration. The other is to calculate the transformation from the 3D measurement system to the patient (i.e., the pose of the anatomical model with respect to the 3D measurement system), which is called image registration. The device calibration is performed offline only once because the 3D display and the 3D measurement system are relatively fixed during surgery. However the image registration suffers from patient's movement, which requires the image registration process to be performed automatically in real time.

In the previous work [3–5,8], an optical tracking system was employed for image registration using a manual marker-based registration method. The involvement of markers will hamper common surgical workflow, and the attachment of markers is either invasive or infeasible in many cases. Automatic markerless image registration is preferable in surgical navigation. In our previous work [10,13], a low-priced stereo camera was employed replacing the Polaris tracking system for the image registration task in oral and maxillofacial surgery. A teeth contour tracking method was proposed to calculate the pose of patient's teeth with respect to the stereo camera without manual intervention. However, this method is a 3D-3D matching method requiring that the organ to be registered have sharp 3D contour features; and these features should be easily reconstructed by the stereo camera three-dimensionally. Such conditions are quite strict, hence limit the applicable scope of the 3D contour tracking method. Furthermore, in the previous method, only silhouette features in the captured image pair are used while other useful visual clues such as image gradients on non-silhouette edges are ignored. The incorporation of these information could improve the accuracy and robustness of image registration.

In this study, we further simplify the hardware for image registration by using a single camera. We present a 3D surgical overlay system with automatic markerless image registration. The image registration is achieved by matching the 2D shape of teeth's mesh model with the intraoperative 2D image captured by the camera. The display-camera calibration is performed by solving a perspective-n-point (PnP) problem (i.e., the estimation of camera's extrinsic parameters given its intrinsic parameters and n-point 3D-2D correspondences). The proposed system enables real-time correct 3D image overlay on patient's head and neck area for surgical visualization in oral and maxillofacial surgery.

2 System Overview

The proposed system consists of a 3D display, a translucent mirror (AR window), a monochrome camera and a workstation for information processing, as shown in Fig. 1(a). The 3D display is composed of a high pixel per inch (ppi) liquid crystal display (LCD) and a hexagonal lens array which is placed in front of the LCD.

The 3D image of a mesh model (e.g., in the form of an STL file) is created using computer generated integral imaging (CGII) techniques and can be projected to a specified location and orientation with respect to the 3D display [12]. The 3D image is further overlaid on the patient by the translucent mirror. Surgeons will see a superimposed 3D image through the AR window to acquire visualized anatomical information. In this study, the goal of the system is to realize intraoperative augmented reality visualization in oral and maxillofacial surgery.

Figure 1(b) shows the involved coordinate systems in the overlay system. We denote by T_C, T_D, T_M the camera, 3D display, and model coordinate systems, respectively. T_D is the world coordinate system in which the 3D image is rendered as described in our CGII rendering algorithm [12]. T_M actually represents the image space where the patient is digitalized (e.g., by CT scanning). To overlay the 3D image correctly on the patient, the transformation from T_D to T_M denoted by $\boldsymbol{T}_D^M$ should be determined. Because we have $\boldsymbol{T}_D^M = \boldsymbol{T}_D^C \boldsymbol{T}_C^M$, this raises two problems: display-camera calibration and image registration.

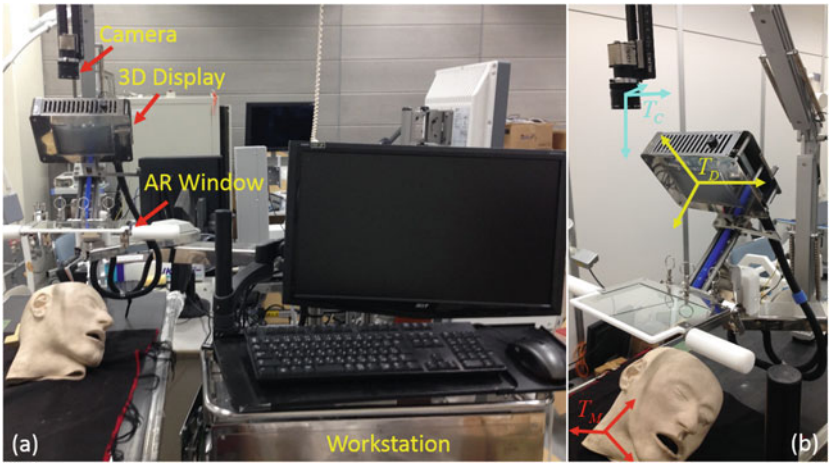


Fig. 1. (a) System overview. (b) Coordinate systems.

3 Display-Camera Calibration

Display-camera calibration is to determine the transformation $\boldsymbol{T}_D^C$, which can be formulated as a camera extrinsic calibration problem. The 3D image of a known geometry in T_D is projected by the 3D display, and the 2D image of the 3D image is captured by the camera through the AR window. In the captured image, the 2D-3D correspondences are established. The pose $(\boldsymbol{R}_C^D, \boldsymbol{t}_C^D)$ of the 3D display[1] with respect to the camera can be estimated by solving

$$\min_{\langle \boldsymbol{R}_C^D, \boldsymbol{t}_C^D \rangle} \sum_{i=1}^{N} ||\boldsymbol{K} \left(\boldsymbol{R}_C^D, \boldsymbol{t}_C^D \right) \boldsymbol{X}_i - \boldsymbol{x}_i||^2 \tag{1}$$

[1] $\det(\boldsymbol{R}_C^D) = -1$ due to the reflection of the AR window.

where $\boldsymbol{X}_i \leftrightarrow \boldsymbol{x}_i, i = 1 \cdots N$ are 3D-2D correspondences in the form of homogeneous coordinates; $\boldsymbol{K}$ is the intrinsic matrix of the camera; $||\boldsymbol{X}_i - \boldsymbol{x}_i||$ denotes the underlying 2D distance between $\boldsymbol{X}_i$ and $\boldsymbol{x}_i$.

Figure 2(a) shows the calibration model which is a 5×5 planar ball array. The captured 2D image through the AR window is shown in Fig. 2(b). The centers of the projected balls are automatically detected using a simple threshold followed by an ellipse fitting. Figure 2(c) shows an example of ball center detection. Equation (1) is well known as a PnP problem which can be easily solved using a nonlinear least squares technique [1]. $\boldsymbol{T}_D^C$ is the inverse of the pose $(\boldsymbol{R}_C^D, \boldsymbol{t}_C^D)$.

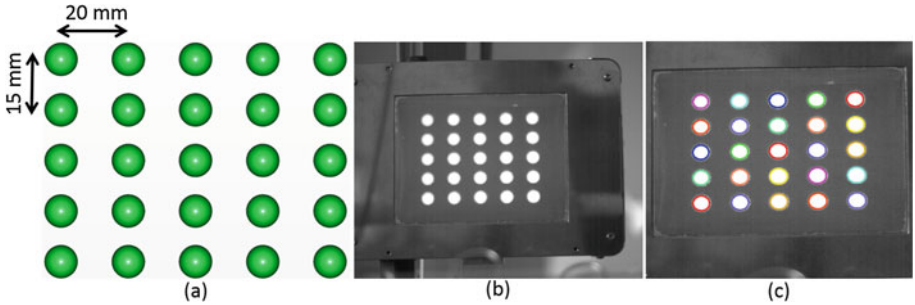


Fig. 2. (a) Calibration model. (b) Captured image of (a). (c) Automatic ball center detection.

4 Image Registration

Image registration is to determine the transformation $\boldsymbol{T}_C^M$. Unlike the fixed spatial relationship during surgery in the device calibration, $\boldsymbol{T}_C^M$ suffers from patient movement and varies when patient's pose is changed. This requires the image registration to be performed automatically in real time. We propose a 3D-2D registration method by matching patient's 3D teeth model (created from preoperative CT data) with the intraoperative 2D camera image based on Ulrich's method [9]. Because teeth are rigid and can be easily exposed to an optical camera, they can serve as "natural markers" for registering anatomical information in the head and neck area.

4.1 Problem Formulation

Our problem is to find the best pose $(\boldsymbol{R}, \boldsymbol{t})$ so that the 2D projection of the 3D model using the projection matrix $\boldsymbol{K}(\boldsymbol{R}, \boldsymbol{t})$ is most consistent with the 2D image. 2D projection views of the 3D model can be rendered using computer graphics API, such as OpenGL. To measure the consistency between the projected model shape E_{2D} and the image $I(x, y)$, we use the following similarity proposed by Steger [7]

$$s(E_{2D}, I) = \frac{1}{N} \sum_{i=1}^{N} \frac{|\langle \nabla I(x_i, y_i), \boldsymbol{d}_i \rangle|}{\|\nabla I(x_i, y_i)\| \cdot \|\boldsymbol{d}_i\|} \tag{2}$$

where $E_{2D} \stackrel{\text{def}}{=} \{x_i, y_i, \boldsymbol{d}_i\}_{i=1}^{N}$ is a set of edge points (x_i, y_i) with associated direction vectors $\boldsymbol{d}_i$ representing the normal direction; $\nabla I(x_i, y_i)$ is the image gradient at (x_i, y_i); $\langle , \rangle$ denotes dot product. The absolute value $|\cdot|$ on the numerator will ignore local contrast polarity change. $s(E_{2D}, I)$ ranges from $[0, 1]$.

Given $(\boldsymbol{R}, \boldsymbol{t})$, the projected 2D shape E_{2D} of the 3D model is extracted as follows. First, the intrinsic matrix $\boldsymbol{K}$ is used to set the view frustum of the virtual camera so that the virtual camera has the same projection geometry as the real camera. Then, the 3D model is rendered into a 3-channel RGB image whose RBG values represent the normal vector on the corresponding surface of the 3D model. Next, the image tensor (2×2 matrix) at each pixel of the RGB image is calculated, whose largest eigen value represents the edge strength of the pixel [6]. The edge strength corresponds to the face angle of the corresponding 3D edge of the model. Subsequently, a threshold is applied to the edge strength to suppress the pixel whose corresponding face angle is below a certain value (e.g., 30°). Finally, non-maximum suppression is performed for edge thinning and the remaining edge pixels with their gradient vectors constitute the 2D shape E_{2D}. Figure 3 shows the extracted 2D shape of a left molar model with the suppression angle of 35°. A straightforward idea is to find optimal $(\boldsymbol{R}, \boldsymbol{t})$ so that (2) is maximized.

4.2 Aspect Graph Based Matching

It is impossible to directly optimize (2) unless the start point is quite near to the true pose. However, we do not have the prior knowledge about the pose of the model. We instead adopt a view-based approach. An aspect graph-based matching method proposed by Ulrich [9] is used for fast 3D-2D pose estimation.

Offline Aspect Graph Building. Views are generated by specifying viewpoints (virtual camera positions) in a spherical coordinate system (SCS) whose origin is set to be the center of 3D model's bounding box. The viewpoint range is specified by $[r_{\min}, r_{\max}]$, $[\varphi_{\min}, \varphi_{\max}]$, and $[\theta_{\min}, \theta_{\max}]$, which is a spherical

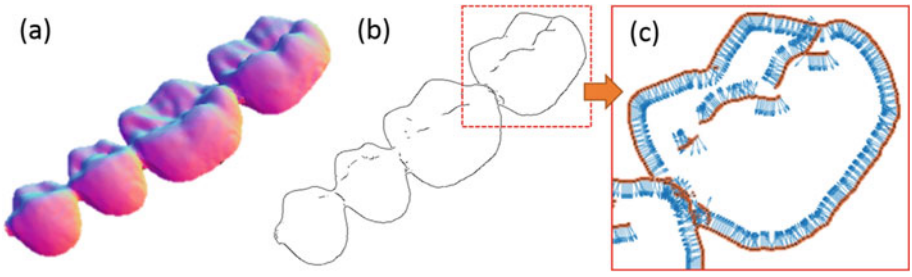


Fig. 3. (a) Rendered RGB image of a molar model. (b) Projected 2D shape of (a). (c) Associated direction vectors.

quadrilateral. r, φ, and θ represent altitude, longitude, and latitude, respectively. The generated views are clustered into aspects according to their mutual similarities calculated using (2) on the overlapped pixels. The aspect in this context means a cluster of views (can be one view) whose mutual similarities are higher than a threshold t_c. A complete-linkage clustering method is used for view clustering. After the clustering is finished, the aspect is downsampled to the next higher image pyramid level and the clustering process is repeated. As a result, we can obtain a hierarchical aspect graph spanning different image levels.

Online 2D Matching. After the hierarchical aspect graph has been built, it is ready to use in the online search phase for pose estimation. We assume the proposed method is used for real-time image registration, in which case the input data is a video stream. A tracking-matching-optimization strategy is proposed for robust and fast pose estimation. The output of the tracking is used to confine the search space at the top image level to a tight bounding box encompassing the object. The tracking-learning-detection (TLD) framework [2] is incorporated for tracking the 2D appearance of the 3D model over a video stream at a higher image level whose resolution is close to 512×512. Let B^n_{target} be the tracked bounding box at the top image level n, denoted by $I^n(x, y)$. All aspects at the top level of the hierarchical aspect graph are examined within B^n_{target} in $I^n(x, y)$. An aspect is represented by its shape features $E_{2D} = \{x_i, y_i, \boldsymbol{d}_i\}_{i=1}^N$. To search for a match of an aspect, E_{2D} is scaled, rotated, and translated by discrete steps as follows

$$x'_i = x_i \sigma \cos\gamma - y_i \sigma \sin\gamma + t_x \tag{3}$$

$$y'_i = x_i \sigma \sin\gamma + y_i \sigma \cos\gamma + t_y \tag{4}$$

where σ is scaling factor; γ is rotation angle; (t_x, t_y) is translation. The similarity between the transformed E_{2D} and the image $I^n(x, y)$ is calculated using (2). Those 2D poses $(\sigma, \gamma, t_x, t_y)$ with resulting similarity exceeding a threshold t_s are stored in a candidate list. These candidates are either refined in the child aspects by searching close neighboring poses, or discarded due to lower similarity than t_s. All candidates are tracked down along the hierarchical level until reaching the bottom image level. The best candidate is considered as the candidate with the highest similarity score. The 3D pose can be recovered from the 2D pose of the best match at the bottom level and the pose of its associated aspect.

4.3 3D Pose Refinement

The accuracy of the obtained 3D pose from matching is usually insufficient due to the discrete step widths when searching for matches. Iterative optimization using an iterative closest point (ICP) algorithm is performed to refine the 3D pose by alternately identifying feature correspondences and estimating the pose. The corresponding 3D point (X_i, Y_i, Z_i) of the shape feature (x_i, y_i) can be recovered

using the z buffer value of OpenGL. The sub-pixel edge point (x'_i, y'_i) near (x_i, y_i) is localized along the direction of $\boldsymbol{d}_i$ using the method proposed in [11]. The refined pose can be calculated by solving the PnP problem $(X_i, Y_i, Z_i) \leftrightarrow (x'_i, y'_i)$. The above procedure is repeated until convergence. Usually, several iterations will lead to satisfactory convergence.

5 Experiments and Results

5.1 Experimental Setting

Figure 4(a) shows the experimental scene. The LCD (6.4 inch) of the 3D display has a resolution of 1024 × 768 pixels with a pixel pitch of 0.13 mm (200 ppi). The micro lens array has lens pitches of 0.89 mm and 1.02 mm in the vertical and horizontal directions, respectively.

A mandibular phantom was created using a 3D printer from real patient's CT data as shown in Fig. 4(b). Fiducial points were made in the front teeth area and molar area of the phantom with known positions in the image (model) space, for accuracy evaluation. The front teeth model and the molar model shown in Fig. 4(c) are used for image registration. Which model should be used depends on the exposed area (front teeth area or molar area).

The camera (UI-3370CP-M-GL, IDS Imaging Development Systems GmbH, Germany) employed in the experiments has a resolution of 2048 × 2048 pixels with maximum frame rate of 80 frames per second (fps). Camera calibration was performed in advance to obtain the intrinsic matrix and remove lens distortion. The computer used in the experiments has an Intel® Core™ i7-4820K CPU (3.7GHz) and a NVIDIA® GeForce GTX TITAN Black GPU. The GPU is used to accelerate the aspect graph building process and the online matching process by parallel computing.

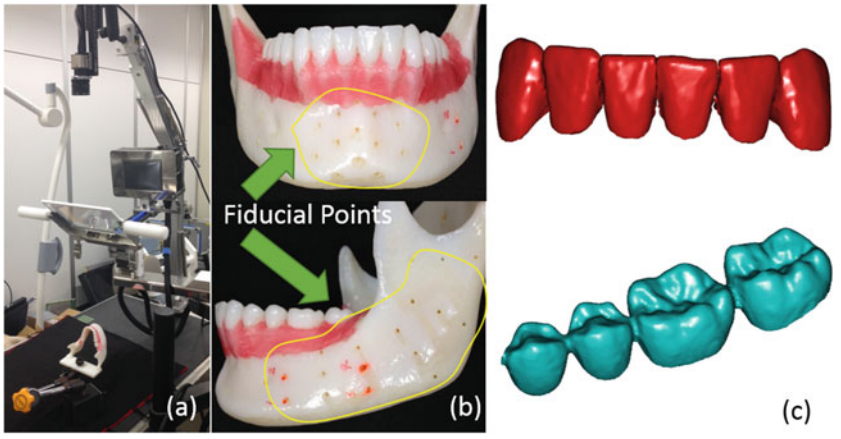


Fig. 4. (a) Experimental scene. (b) Mandibular phantom with fiducial points. (c) Teeth models for image registration.

5.2 Display-Camera Calibration

Figure 5 shows the automatic processing flow of the display-camera calibration. The original image was smoothed using a Gaussian filter. White spots were segmented out using a threshold followed by classifications according to their roundness and areas. The contours of the extracted spots were approximated by ellipses whose centers were used for PnP estimation. The PnP estimation yielded a geometric estimation error of 1.9 pixels.

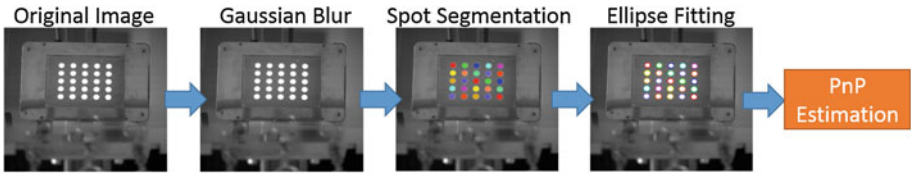


Fig. 5. Processing flow of display-camera calibration.

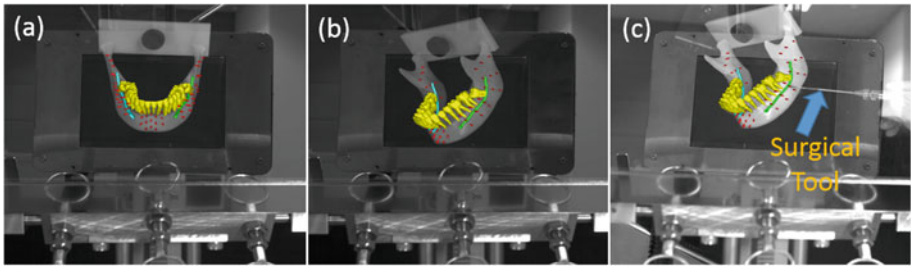


Fig. 6. Image registration results using (a) front teeth model (b) left molar model. (c) Target registration evaluation.

5.3 Image Registration Evaluation

Image registration was performed by matching the model with the video stream captured through the AR window. The distance between the camera and the phantom was approximately 710 mm. The registration process took approximately 0.2 s yielding an update frame rate of 5 fps. Figure 6(a) and (b) show the image registration results using the front teeth model and the left molar model (see Fig. 4(c)), respectively, with critical structures (tooth roots and nerve channels) and fiducial points (in red) overlaid on camera's view.

Target registration errors (TREs) on the fiducial points (see Fig. 4(b)) were calculated to evaluate the registration accuracy as follows. A surgical tool (dental driller) was used to approach individual fiducial points on the phantom under the guidance of the virtually overlaid fiducial points on camera's view. The physical distance between the indicated position and the real position of a fiducial point

on the phantom was measured as the error on that fiducial point. The average error distance in an evaluation area was calculated as the TRE in that area. For TRE calculation in the front teeth (left molar) area, the front teeth (left molar) model was used for the image registration. Figure 6(c) shows the accuracy evaluation process. The accuracy evaluation results yielded TREs of 0.8 mm in the front teeth area (15 points) and 1.1 mm in the molar area (18 points).

5.4 3D Surgical Overlay

After display-camera calibration and image registration, the necessary spatial information for 3D display has become available. Figure 7 shows the 3D overlay of the tooth roots and nerve channels, observed through the AR window. The visualized information could be used to guide surgical operation.

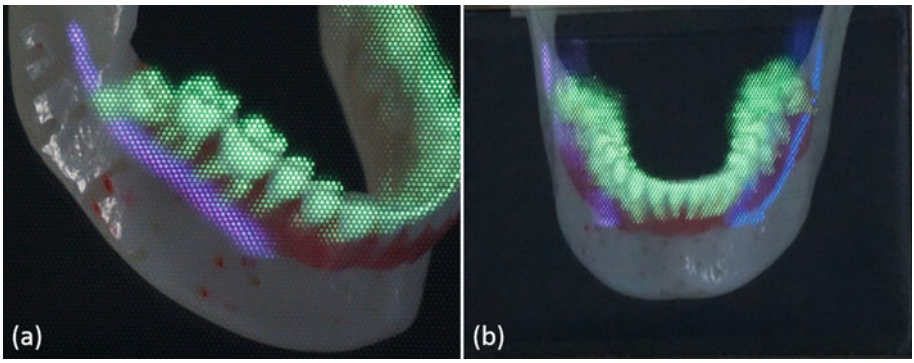


Fig. 7. 3D surgical overlay by (a) molar matching and (b) front teeth matching.

6 Conclusion

This paper presents a 3D surgical overlay system with automatic markerless image registration using a single camera. Teeth are rigid and easy to be exposed to a camera, making it possible to match a teeth model with an intraoperative camera image for the image registration task. The registered 3D images representing anatomical structures are superimposed on the patient via a translucent mirror for augmented reality surgical visualization. In this study, the application in dental surgery was demonstrated using our proposed system. Given the fact that the maxillary teeth are fixed with the skull, the proposed method may also be used for surgical navigation in the craniofacial region. In that case, the error compensation in the area far from the registration features can be a challenging work.

Acknowledgment. This work was supported in part by JSPS Grant-in-Aid for Scientific Research on Innovative Areas (Multidisciplinary Computational Anatomy) JSPS KAKENHI Grant Number 26108008, and a Translational Research Network Program grant from the Ministry of Education, Culture, Sports, Science and Technology, Japan. H. Liao was supported in part by National Natural Science Foundation of China (Grant No. 81427803, 61361160417, 81271735).

References

1. Hartley, R.I., Zisserman, A.: Multiple View Geometry in Computer Vision, 2nd edn. Cambridge University Press, Cambridge (2004). ISBN: 0521540518
2. Kalal, Z., Mikolajczyk, K., Matas, J.: Tracking-learning-detection. IEEE Trans. Pattern Anal. Mach. Intell. **34**(7), 1409–1422 (2012)
3. Liao, H., Hata, N., Nakajima, S., Iwahara, M., Sakuma, I., Dohi, T.: Surgical navigation by autostereoscopic image overlay of integral videography. IEEE Trans. Inf. Technol. Biomed. **8**(2), 114–121 (2004)
4. Liao, H., Inomata, T., Sakuma, I., Dohi, T.: 3-D augmented reality for MRI-guided surgery using integral videography autostereoscopic image overlay. IEEE Trans. Biomed. Eng. **57**(6), 1476–1486 (2010)
5. Liao, H., Ishihara, H., Tran, H.H., Masamune, K., Sakuma, I., Dohi, T.: Precision-guided surgical navigation system using laser guidance and 3D autostereoscopic image overlay. Comput. Med. Imaging Graph. **34**(1), 46–54 (2010)
6. Sebastian, T., Klein, P., Kimia, B.: On aligning curves. IEEE Trans. Pattern Anal. Mach. Intell. **25**(1), 116–125 (2003)
7. Steger, C.: Occlusion, clutter, and illumination invariant object recognition (2002)
8. Tran, H., Suenaga, H., Kuwana, K., Masamune, K., Dohi, T., Nakajima, S., Liao, H.: Augmented reality system for oral surgery using 3D auto stereoscopic visualization. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part I. LNCS, vol. 6891, pp. 81–88. Springer, Heidelberg (2011)
9. Ulrich, M., Wiedemann, C., Steger, C.: Combining scale-space and similarity-based aspect graphs for fast 3D object recognition. IEEE Trans. Pattern Anal. Mach. Intell. **34**(10), 1902–1914 (2012)
10. Wang, J., Suenaga, H., Hoshi, K., Yang, L., Kobayashi, E., Sakuma, I., Liao, H.: Augmented reality navigation with automatic marker-free image registration using 3D image overlay for dental surgery. IEEE Trans. Biomed. Eng. **61**(4), 1295–1304 (2014)
11. Wang, J., Kobayashi, E., Sakuma, I.: Coarse-to-fine dot array marker detection with accurate edge localization for stereo visual tracking. Biomed. Sig. Process. **15**, 49–59 (2015)
12. Wang, J., Suenaga, H., Liao, H., Hoshi, K., Yang, L., Kobayashi, E., Sakuma, I.: Real-time computer-generated integral imaging and 3D image calibration for augmented reality surgical navigation. Comput. Med. Imaging Graph. **40**, 147–159 (2015)
13. Wang, J., Suenaga, H., Yang, L., Liao, H., Kobayashi, E., Takato, T., Sakuma, I.: Real-time marker-free patient registration and image-based navigation using stereovision for dental surgery. In: Liao, H., Linte, C.A., Masamune, K., Peters, T.M., Zheng, G. (eds.) MIAR 2013 and AE-CAI 2013. LNCS, vol. 8090, pp. 9–18. Springer, Heidelberg (2013)

Simultaneous Estimation of Feature Correspondence and Stereo Object Pose with Application to Ultrasound Augmented Robotic Laparoscopy

Uditha L. Jayarathne[1,2(✉)], Xiongbiao Luo[2], Elvis C.S. Chen[1,2], and Terry M. Peters[1,2]

[1] Biomedical Engineering Graduate Program, Western University, London, Canada
[2] Robarts Research Institute, Western University, London, Canada
{ujayarat,xluo,chene,tpeters}@robarts.ca

Abstract. In-situ visualization of ultrasound in robot-assisted surgery requires robust, real-time computation of the pose of the intra-corporeal ultrasound (US) probe with respect to the stereo-laparoscopic camera. Image based, intrinsic methods of computing this relative pose need to overcome challenges due to irregular illumination, partial feature occlusion and clutter that are unavoidable in practical robotic-laparoscopy. In this paper, we extend a state-of-the-art simultaneous monocular pose and correspondence estimation framework to a stereo imaging model. The method is robust to partial feature occlusion and clutter, and does not require explicit feature matching. Through exhaustive experiments, we demonstrate that in terms of accuracy, the proposed method outperforms the conventional stereo pose estimation approach and the state-of-the-art monocular camera-based method. Both quantitative and qualitative results are presented.

1 Introduction

Laparoscopic ultrasound (LUS) is an essential intraprocedural imaging modality that provides real-time imaging of hidden structures in many robot-assisted laparoscopic interventions. During US guided tasks, US video is typically visualized side-by-side with that from the laparoscopic camera, requiring the surgeon to mentally fuse information from one modality to the other. This cognitively mediated information integration strategy is known to cause excessive cognitive load, leading to surgical errors [1]. In-situ visualization avoids such complications, but requires robust, real-time tracking of the US probe with respect to the laparoscopic camera in six degrees of freedom (6DoF).

To solve this relative pose problem, the use of extrinsic tracking systems has been demonstrated with magnetic [2], optical [3], magneto-optic hybrid [4] and robot-kinematic-based [5] systems. In addition to high cost, these extrinsic tracking systems disturb the current operating room (OR) workflow and demand impractical tracking environments (magnetically clean environment for

C.A. Linte et al. (Eds.): AE-CAI 2015, LNCS 9365, pp. 134–144, 2015.
DOI: 10.1007/978-3-319-24601-7_14

magnetic tracking and optical line of sight for optical tracking). As a solution, image-based intrinsic tracking methods [6] have been introduced that can be seamlessly integrated into the existing OR workflow. However, these methods are challenging primarily due to irregular illumination, and feature association issues in the presence of occlusion and clutter.

In order to explicitly solve the feature association problem, appearance based descriptors (e.g. SIFT, SURF) are often used. These descriptors tend to fail in the presence of irregular illumination in addition to their high localization uncertainty [7]. Hence they are not appropriate for practical, accuracy-demanding laparoscopic applications. Without relying on appearance cues, Jayarathne et al. [6] proposed an Extended Kalman Filter (EKF)-based algorithm to jointly solve for the 3D-2D feature correspondence and 6DoF object pose using a monocular camera. However, to date no attempt has been made to extend this method to stereo (and general multi-view case) to exploit additional information in multiple views which may improve depth estimates and tracking range.

The conventional approach to stereo (and multi-view) pose estimation is to explicitly establish feature correspondence across views, triangulate the matched features and extract the pose parameters by fitting the 3D object model to the triangulated points with an Iterated Closest Point (ICP)-based algorithm [8]. In the presence of feature occlusion due to the curvature of the probe, blood or other surgical tools, and clutter due to imperfections of the feature detector, these methods often fail to provide accurate pose estimates. Moreover, feature matching based on merely epipolar constraints could result in spurious matching which often leads to erroneous estimates. Typically RANdom SAmple Consensus (RANSAC) [9] algorithm based post processing techniques are employed to remedy this, but such methods affect the run-time performance of the algorithm.

In this paper, we extend the Kalman Filter based, joint pose and correspondence estimation framework [6] to a stereo imaging model, but it can readily be applied to the general multi-view case. We demonstrate through exhaustive experiments that the proposed method outperforms the conventional triangulation based method, and that its pose estimates are more accurate compared to the single camera EKF-based method. Section 2 describes the proposed method and experiments in detail, Sect. 3 summarizes the results and Sect. 4 concludes the paper with a detailed discussion and offers some insight into future work.

2 Materials and Methods

2.1 Experimental Setup

Camera Geometric Calibration: In all our experiments we used an Olympus Stereoscopic camera from the *daVinci S* surgical system. The video feed from the left and the right camera was captured at 640×480 resolution. The well-known generic camera calibration method proposed in [10] was employed with a checkerboard calibration pattern to determine the intrinsic parameters of each camera with corresponding distortion parameters. The fundamental matrix of the stereo-pair was derived with a method based on least median of squares [11].

(a) (b) (c)

Fig. 1. (a) A fiducial pattern is glued on the back surface of the 3D-printed cover for the Ultrasonix US probe. (b) Optically tracked validation tool with a mock-probe rigidly placed in middle. (c) US calibration tool. The 3D-printed straw holder is tracked with a checkerboard pattern glued on its flat surface.

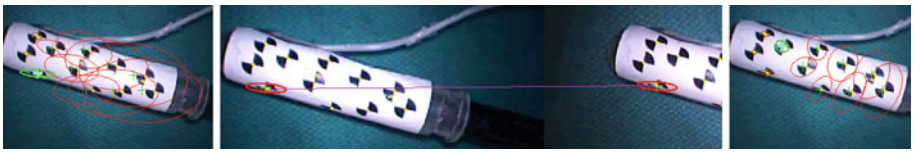
Features and Feature Detection: Inspired by the idea of using a curved "X-corner" pattern for accurate 6DoF tracking of a clinical LUS probe presented in [6], an "X-corner" pattern was glued on the curved back surface of a 3D-printed probe-tip cover (Fig. 1(a)) for the LUS probe (LAP9-4/38, Ultrasonix, Canada). The 3D position of the "X-corners" were measured accurately with a measurement microscope (STM6-LM, Olympus, Japan), and a local coordinate system was defined on the marker space with one corner arbitrarily chosen as the origin. The 3D point-set thus defined serves as the model. The "X-corners" were detected with an efficient algorithm [12] in each left and right camera frame, and were further refined to sub-pixel accuracy.

2.2 Joint Estimation Framework

Alleviating the issues in the conventional triangulation-based pose estimation approach, we solve both the feature correspondence problem and the pose problem simultaneously in an efficient Bayesian framework. This allows natural integration of priors on the pose to drastically reduce the search space, resulting in real-time performances. In addition, the framework provides easy access to uncertainty information of the estimation parameters which could be useful in analyzing tracking performance.

Pose Parameterization and Priors: The pose of the probe was parameterized as a 6D vector: three parameters representing 3D rotation[1] and three 3D translation. To sample this six dimensional pose space, an optically tracked mock probe with a fiducial pattern (Fig. 1(b)) was moved freely within the camera field of view, while the camera itself was being tracked. The hand-eye calibration, and the transform between the fiducial pattern and the optical tracking dynamic reference body (DRB) was determined (see Experiments section for details), and the pose of the fiducial pattern with respect to the camera image was saved. Finally a Gaussian Mixture Model (GMM) was learned [13] over these collected pose

[1] In Rodrigues representation.



(a) (b) (c)

Fig. 2. An instance of the algorithm. In (a) and (c) only the left image is shown due to space limitations: (a) Model points are projected to the image space (green + marker) with their covariance (in red ellipses while the bold green ellipse represents the one currently being evaluated) (b) Within the ellipse currently being evaluated, the correspondence between the detected features (yellow dots) across images is determined based on minimum Sampsons distance (c) image space uncertainty drops after the UKF update. This is indicated by the shrinkage of the search ellipses (Color figure online).

samples providing pose priors for initialization. Once initialized, the Gaussian prior for the subsequent frame is derived based on a motion model of the probe.

Joint Estimation: Let $x = [r_1, r_2, r_3, t_x, t_y, t_z]$ be the pose, P_x be its covariance matrix and $Proj_K(M_i, x)$ be the operator that projects the model point M_i onto the K^{th} camera with the pose x. We begin with the mean pose and the covariance of the GMM component with highest weighting.

Each model point M_i is projected to the stereo-image space with the pose x with its image space locations, y_i^L and y_i^R respectively in left and the right camera, and their covariance, $P_{y_i}^L$ and $P_{y_i}^R$ respectively, estimated with the Unscented Transform [14]. The Unscented Transform avoids complex Jacobian computation in the EKF-based approach [6] and is superior in handling non-linearities in the stereo-imaging model. Feature points in each image (left/right) with less than a predefined *Mahalanobis* distance from y_i^L (and y_i^R) are then considered to be possible matches for the projected model point. The ellipsoid defined by this *Mahalanobis* distance defines the search region for model-to-feature correspondence (Fig. 2(a)). In all our experiments we use *Mahalanobis* distance of 3, corresponding to 99 % confidence interval.

For each feature point lying inside the elliptical search region in the left camera Sampsons distance [15] is computed to each feature point in the right camera lying inside the corresponding search ellipse (Fig. 2(b)). Feature pairs with less than a predefined Sampsons distance threshold are then considered to be potential left-right matched pairs for the model point. Assuming that one of these matching pairs is the correct model-to-feature correspondence, we update the pose and its covariance using Unscented Kalman filter (UKF) state update equations [14] (For better numerical stability we use the strategy presented in Square-Root Kalman (SRUKF) [16] filter whose practical run-time complexity is better than that of the standard UKF). This update reduces the covariance of the pose further constraining the search regions. This is illustrated by the shrinkage of the search ellipses in Fig. 2(c). Model points are projected back

again with the updated pose and the covariance, and possible matching pairs determined as described above. After three such updates the evolution of the pose (and the covariance) becomes negligible which confirms that three point correspondences are adequate to determine the 6DoF pose uniquely. At this point, the model points are projected into the left and right cameras and the following error function is computed.

$$\begin{aligned} Error(x) &= \sum\nolimits_{i=1,\ldots,N} |Proj_L(M_i, x) - U_L| \\ &+ \sum\nolimits_{i=1,\ldots,N} |Proj_R(M_i, x) - U_R| + \gamma |NotDetected| \end{aligned} \tag{1}$$

where U_L and U_R are feature points in the left and right cameras respectively with closest euclidean distance to the projection of model point M_i, $|NotDetected|$ is the cardinality of the not detected model points, and γ is a tuning parameter. Feature points, lying at a distance greater than the variance of the sensor noise from the closest projected model point, are considered to be outliers.

The algorithm iterates through all the model-feature matched-pairs, and through all the GMM components until the error in Eq. (1) drops below a pre-defined threshold. Once converged, the updated pose is very close to the correct one. Note that at this point we know the correct model-feature correspondence as well as the correct left-right correspondence. Finally, this pose is further refined by applying SRUKF updates per each model-feature correspondence. This refined pose is considered to be the correct pose for the processed frame.

The prior pose and covariance for the next frame is given by an SRUKF state prediction step with the identity process model. However, with sophisticated process models that closely simulate the probe motion, better tracking performances can be expected.

2.3 Experiments

To evaluate the performance of the proposed two-view pose estimation framework, several experiments were performed. A commercial optical tracking system (Spectra, NDI, Canada) provided ground truth poses by tracking the stereo-endoscope and the special assembly with the 3D printed probe mock-up which will be termed *the validation tool* from here onwards (Fig. 1(b)). To be able to compare the image-based estimates with the ground truth, we estimated the hand-eye calibration matrix with a globally optimal algorithm [17] while the transformation between the fiducial pattern and the optically tracked DRB was determined by a robust method [18]. Pose estimates were computed offline on a computer with Intel Core i7 (3.4 GHz) 64-bit CPU and 16 GB of RAM running Microsoft Windows 7.

(a) (b) (c)

Fig. 3. (a) The CAD design for the 3D-printed straw-holder used in US calibration. (b) Left camera image showing an instance in data acquisition for US calibration. (c) US image of the hollow straw

Comparison to Ground Truth: The validation tool was moved freely in front of the camera while the endoscope and the validation tool were being tracked by the optical tracking system. A stereo video was saved and the pose estimates were computed offline using the proposed method implemented in Matlab. The EKF-based method presented in [6] was implemented in Matlab and pose estimates were computed using the left camera frames of the saved stereo video. To avoid feature matching errors in the conventional triangulation based approach, the feature matches determined by the proposed method were used. The matched features were triangulated [15] and the pose parameters were extracted by aligning the model with the triangulated points with standard Procrustes analysis.

Ultrasound Calibration and Image Overlay: To overlay US images in endoscopic video, the US calibration transform was determined with a method that formulates the problem as a point-to-line registration [19]. A plastic straw was imaged while it was tracked with a checkerboard pattern glued to its 3D-printed holder (Figs. 1(c) and 3(b)) designed such that the central axis of the straw is known relative to the coordinate system defined on the pattern (Fig. 3(a)). The pose of the probe was determined with the proposed method. The centroids of the elliptical representations of the ultrasound images of the straw (Fig. 3(c)) were determined manually in each image, and the end points of the tracked straw were then used as inputs to the Anisotropic-Scaled ICP (AS-ICP) [20] algorithm, where the point-to-line registration transform in form of anisotropic scales, followed by rotation and translation is the US calibration we seek.

To qualitatively evaluate the performance, a hollow plastic straw submerged in a water-bath was imaged with the probe tracked by the proposed method. Pose estimates with US calibration transform were then used to overlay US images in the stereoscopic video.

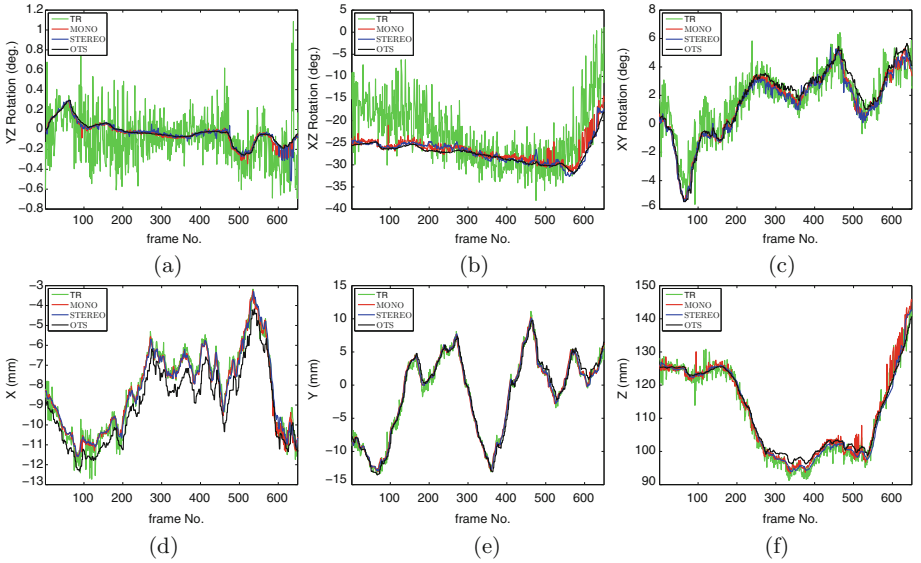


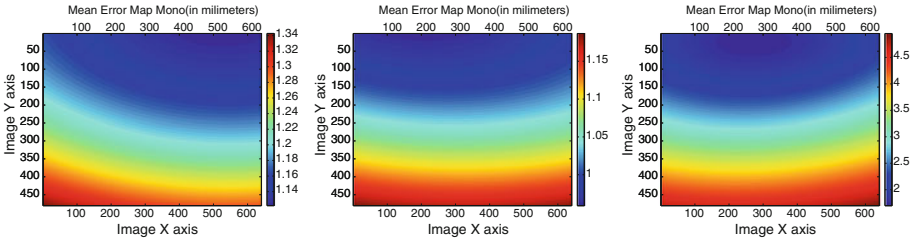
Fig. 4. Estimates computed by **mono** (in red), **stereo** (in blue) and **tr** (in green) methods. The ground truth is in black. (a) Rotation in YZ-plane. (b) Rotation in XZ-plane. (c) Rotation in XY-plane. (d) Translation along X-axis. (e) Translation along Y-axis. (f) Translation along Z-axis (Color figure online).

3 Results

Poses computed by the method presented in [6], the proposed method and the triangulation-based method, referred to here as **mono**, **stereo** and **tr** respectively, were compared to the optical tracker based ground truth, referred to as **ots**. Over a total of 650 frames, the root mean square (RMS) translational error was 2.0 ± 1.4 mm, 1.6 ± 0.9 mm and 2.4 ± 1.3 mm for **mono**, **stereo** and **tr** respectively. The results are summarized in Fig. 4. The average computation time for the EKF based method in [6] was 15.7 ms, while that for the proposed method was 38.3 ms.

Parameter estimates, particularly those related to rotation and the translation along the Z-axis, resulting from **tr** are jittery, and deviate significantly from the ground truth. Even though estimates computed by **mono** follow the ground truth fairly closely, significant jitter and errors are observed in the rotational estimates about the Y-axis and translational estimates along the Z-axis of the camera (Fig. 4(b) and (f)). The estimates computed by the proposed method follow the ground truth very closely and smoothly.

A total of 20 images, captured at an imaging depth of 4 cm, were used for ultrasound calibration, resulting in Fiducial Registration Error (FRE) of 0.57 mm. The estimated US calibration transform together with the probe poses estimated by the proposed method were used to overlay US images in the stereoscopic video. Figure 6(a) shows a snap-shot of the overlaid video corresponding

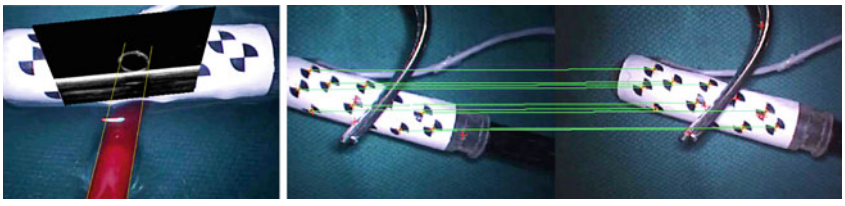


(a) (b) (c)

Fig. 5. The mean TRE maps for (a) **mono** method, (b) **stereo** method and (c) **tr** method. The X and Y axes correspond to that of the US image while the origin of the fiducial pattern is close to the top-left corner. Errors are in millimeters. Note that the scale in the error bar is different in each image

to the left camera. Note that the extended edges of the straw perfectly intersect with its US image, providing a qualitative measure of the accuracy of the estimates.

Considering 3D points corresponding to each US pixel as targets, the Target Registration Error (TRE) was computed for **mono**, **stereo** and **tr** methods, using the estimated US calibration matrix to transform US pixel locations to the coordinate system defined on the fiducial pattern. The mean TRE maps for each method are shown in Fig. 5. The **tr** method demonstrates the worst TRE (mean ranging from 1.6 mm – 4.9 mm), while the TRE resulted in **stereo** demonstrates the best (mean ranging from 0.9 mm – 1.1 mm). Even though it is minor, an improvement in the mean TRE is observed in **stereo** compared to that in **mono** (compare range of mean TRE in **mono**, 1.1 mm – 1.3 mm, to that in **stereo**).



(a) (b)

Fig. 6. (a) US image overlaid onto the left camera image using the pose estimated by the proposed method and the estimated US calibration matrix. Note that the extended edges of the straw (in red) perfectly intersect with its US image. The solid white line in the US image corresponds to the strong reflection of sound waves from the bottom of the water-bath. (b) Stereo image showing that the proposed method has accurately determined the left-right feature correspondence even in the presence of clutter and partial feature occlusion (Color figure online).

Figure 6(b) qualitatively illustrates the robustness of the algorithm to clutter and partial feature occlusion. The left-right feature correspondence has accurately been determined even though the fiducial pattern is partially occluded by a surgical tool.

4 Discussion and Conclusion

In this paper, an efficient image based method for stereo pose estimation is presented. Extending the state-of-the-art method [6] to stereo-imaging model, it simultaneously solves for both the feature correspondence and the object pose while maintaining robustness to feature occlusion and clutter. The method is compared to the conventional triangulation-based approach using an optical tracking system as the ground truth. In addition, the performance of the proposed algorithm is compared to the state-of-the-art monocular image-based method to demonstrate the improvements due to the use of multiple views for pose estimation.

The proposed method produces more accurate and stable pose estimates compared to the conventional triangulation-based method and the state-of-the-art single-view-based method. The improvement in precision and accuracy is significant in the rotational estimates and translation along the depth direction. This may be explained by the wealth of depth information in multiple views. The smooth evolution of the estimates computed by the proposed and the monocular-image-based methods is a result of the inbuilt Kalman Filter. The mean TRE maps further confirms the improvements in tracking resulted in the proposed method compared to the others. Note that the imperfect registration between the coordinate system of the optical tracking system and that centered in the imaging plane of the camera has some contribution to the TRE computation. Future work will look into methods to minimize this error contribution to achieve more accurate estimates of the TRE.

The proposed method uses Unscented Transforms to propagate uncertainty from pose space to the measurement (image) space, while the single-view-based method uses Jacobians that can be computed efficiently. This could be the cause of the added computational burden resulting in increased average computation time in the proposed method. However, we believe that real-time performance can still be achieved with efficient C/C++ implementation of the algorithm, which will be a part of our future work.

Acknowledgements. This work was supported by the Canadian Institutes of Health Research, the Collaborative Research and Training Experience (CREATE) Program of the National Science and Engineering Research Council of Canada, and the Canadian Foundation for Innovation. Authors would like to thank John Moore for his invaluable assistance in 3D printing.

References

1. Wu, B., Klatzky, R.L., Shelton, D., Stetten, G.D.: Psychophysical evaluation of in-situ ultrasound visualization. IEEE Trans. Vis. Comput. Graph. **11**(6), 684–693 (2005)
2. Cheung, C.L., Wedlake, C., Moore, J., Pautler, S.E., Peters, T.M.: Fused video and ultrasound images for minimally invasive partial nephrectomy: a phantom study. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part III. LNCS, vol. 6363, pp. 408–415. Springer, Heidelberg (2010)
3. Kang, X., Azizian, M., Wilson, E., Wu, K., Martin, A.D., Kane, T.D., Peters, C.A., Cleary, K., Shekhar, R.: Stereoscopic augmented reality for laparoscopic surgery. Surg. Endosc. **28**, 1–9 (2014)
4. Feuerstein, M., Reichl, T., Vogel, J., Schneider, A., Feussner, H., Navab, N.: Magneto-optic tracking of a flexible laparoscopic ultrasound transducer for laparoscope augmentation. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part I. LNCS, vol. 4791, pp. 458–466. Springer, Heidelberg (2007)
5. Leven, J., et al.: DaVinci canvas: a telerobotic surgical system with integrated, robot-assisted, laparoscopic ultrasound capability. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3749, pp. 811–818. Springer, Heidelberg (2005)
6. Jayarathne, U.L., McLeod, A.J., Peters, T.M., Chen, E.C.S.: Robust intraoperative US probe tracking using a monocular endoscopic camera. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013, Part III. LNCS, vol. 8151, pp. 363–370. Springer, Heidelberg (2013)
7. Zeisl, B., Georgel, P.F., Schweiger, F., Steinbach, E., Navab, N.: Estimation of location uncertainty for scale invariant feature points. In: Procedings of British Machine Vision Conference, pp. 1–12 (2009)
8. Choi, C., Baek, S.-M., Lee, S.: Real-time 3D object pose estimation and tracking for natural landmark based visual servo. In: IEEE International Conference on Intelligent Robots and Systems (2008)
9. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM **24**, 381–395 (1981)
10. Zhang, Z., Member, S.: A flexible new technique for camera calibration. IEEE Trans. Pattern Anal. Mach. Intell. **22**(11), 1330–1334 (2000)
11. Kaehler, A., Bradski, G.: Learning OpenCV: Computer Vision with the Opencv Library. O'Reilly, Sebastopol (2008)
12. Bennett, S., Lasenby, J.: ChESS: quick and robust detection of chess-board features. Comput. Vis. Image Underst. **118**, 197–210 (2014)
13. Figueiredo, M., Jain, A.: Unsupervised learning of finite mixture models. IEEE Trans. Pattern Anal. Mach. Intell. **24**, 381–396 (2002)
14. Wan, E., Van Der Merwe, R.: The unscented kalman filter for nonlinear estimation. In: IEEE Adaptive Systems for Signal Processing, Communications, and Control Symposium, pp. 153–158 (2000)
15. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press, New York (2003)
16. Van der Merwe, R., Wan, E.: The square-root unscented Kalman filter for state and parameter-estimation. In: 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings, pp. 3461–3464 (2001)
17. Heller, J., Havlena, M., Pajdla, T.: A branch-and-bound algorithm for globally optimal hand-eye calibration. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1608–1615 (2012)

18. Shah, M.: Comparing two sets of corresponding six degree of freedom data. Comput. Vis. Image Underst. **115**, 1355–1362 (2011)
19. Ameri, G., McLeod, A.J., Baxter, J.S.H., Chen, E.C.S., Peters, T.M.: Line fiducial material and thickness considerations for ultrasound calibration. In: Proceedings of SPIE Medical Imaging (2015). 941529–941529-9
20. Chen, E., McLeod, A., Baxter, J., Peters, T.: Registration of 3d shapes under anisotropic scaling. Int. J. Comput. Assist. Radiol. Surg. **10**, 1–12 (2015)

Patient Adapted Augmented Reality System for Real-Time Echocardiographic Applications

Gabriel Kiss(✉), Cameron Lowell Palmer, and Hans Torp

Department of Circulation and Medical Imaging and MI Lab,
Norwegian University of Science and Technology, Trondheim, Norway
gabriel.kiss@ntnu.no
https://www.ntnu.edu/employees/gabriel.kiss

Abstract. When compared to other imaging modalities, the acquisition and interpretation of ultrasound data is challenging for new users. The main aim of this work is to investigate whether augmented reality can provide a patient specific correspondence between the ultrasound image and a virtual anatomic model of the heart and thus simplify the live echocardiographic scanning process.

The proposed augmented view contains the image of the scene (acquired with a tablet's camera), a generic heart model and the ultrasound image plane. The position and orientation of the ultrasound probe and patient are tracked with the Vuforia framework, using fiducial markers that are visible on the tablet's camera image. A customized renderer for augmentation purposes was implemented in OpenGL ES. Data streaming from a high end ultrasound scanner to the tablet device was implemented and the position of the heart model is continuously updated to match the ultrasound data by tracking and matching a set of anatomic landmarks. The prototype was tested in the echo laboratory and real-time performance was achievable as no significant lag between the scanner image and the one presented on the tablet was experienced. Furthermore the suitability of Vuforia for fiducial tracking was evaluated and was deemed sufficiently accurate for this application. The presented prototype was tested by an experienced echocardiographer and was considered beneficial for both teaching purposes and as a help tool for inexperienced ultrasound users.

Keywords: Ultrasound · Echocardiography · Augmented reality · Vuforia

1 Introduction

The acquisition and interpretation of ultrasound images remains a great challenge for new users [6], a typical learning process taking between 6 and 12 months. The teaching process should eventually lead to the development of specific cognitive and motor skills, required for good quality image acquisition and interpretation. Since ultrasound image acquisition is operator dependent its successful

C.A. Linte et al. (Eds.): AE-CAI 2015, LNCS 9365, pp. 145–154, 2015.
DOI: 10.1007/978-3-319-24601-7_15

utilization is highly dependent on the experience of the examiner in contrast to computed tomography or magnetic resonance imaging which are less operator dependent. The late 1990's have marked the introduction of the first portable ultrasound units for medical purposes. This poses additional challenges since typically the portable devices are aimed at non-experts. Thus ensuring that a user has a sufficient degree of training, aiding the user during image acquisition as well as making sure that the acquired image has sufficient image quality is of importance for a successful examination.

Augmented reality is an active field of research in the context of medical procedures [5]. The Sonic Flashlight introduced by Stetten et al. [9,10] uses a half-silvered mirror to depict ultrasound images directly on the patient's body. Bajura et al. [1] used a see-through head-mounted display to integrate ultrasound images into a three-dimensional scene. State et al. [8] employed a similar setup for ultrasound-guided biopsies. Studeli et al. [11] presented information from pre-operative CT images combined with live ultrasound data during radio-frequency ablation of liver tumors. However, most of these systems require complex technical setups and typically use head-mounted displays [4] or projector-based setups [2], which are not well suited for bedside examinations. In addition, many of the mentioned tools are aimed at a very specific application and are to be used by highly experienced clinicians, which limits their usability. Systems that minimally interfere with the clinical workflow and are able to provide visual guidance during the echocardiographic image acquisition process have been recently proposed. In [3] a prototype system that employs tablet technology for display purposes is presented. However, the presented approach required an external server for processing purposes, while the visual compositing requires a manual alignment of the patient to the virtual scene. Once this alignment is made no corrections are attempted for the duration of the scan. Moreover, the alignment step has to be repeated for each new scanning.

The main aim of our work is to develop and test an augmented reality based system that will enhance and simplify the echocardiographic acquisition process as well as help non-experienced users during teaching and acquisition. Furthermore, we aim to implement most of the augmentation pipeline on a tablet device and as such address challenges such as affordability, maintenance (use of easy to replace off the shelf hardware) and portability. Additionally, by showing the augmented view on a tablet placed near the patient an in-situ visualization that does not interfere with the clinical workflow is achieved.

2 Methods

Figure 1 presents an overview of our system. A continuous stream of 3D ultrasound data is recorded with a high end ultrasound scanner. A deformable model is fitted to the data in real-time and the position of a set of anatomic landmarks is detected and subsequently tracked in time (Sect. 2.1). A high level heart model triangle mesh is deformed and aligned to the ultrasound data (Sect. 2.2). Simultaneously, the position of both probe and patient is also computed by detecting

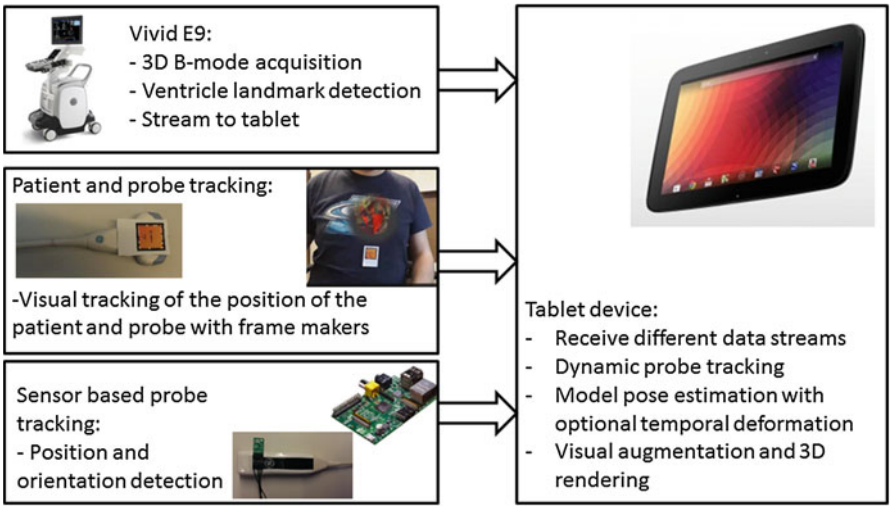


Fig. 1. System overview of the augmented reality prototype. Ultrasound data with annotated anatomic landmarks, visual tracking and optionally the values obtained from a position sensor can be sent to the tablet device.

known frame markers in the image scene (Sect. 2.3). Finally, the augmented view is generated on the tablet device (Sect. 2.4). Practical implementation details are given in Sect. 2.5.

2.1 Heart Detection in the Ultrasound Data

The relative position of the patient's heart with regards to the ultrasound probe can be determined by fitting a deformable model to the 3D ultrasound volumes. The approach proposed by Orderud et al. is employed [7], in which a Kalman based tracker is used. The measurement vector for the Kalman filter is a set of edge detections equally spread on the surface of the left ventricle, whereas the state vector consists of a set of control points defining the shape of a Doo-Sabin subdivision model. Figure 2 shows the principle of the Kalman tracker with the following equations giving the measurement update:

$$\hat{x}_k = \overline{x}_k + \hat{P}_k H^T R^{-1} v_k \tag{1}$$

$$\hat{P}_k^{-1} = \overline{P}_k^{-1} + H^T R^{-1} H \tag{2}$$

with $\hat{x}_k$ the updated state estimate, H includes the edge measurements, R represents the measurement covariance matrix, $\hat{P}$ the updated error covariance matrix and $\hat{P}_k H^T R^{-1}$ the Kalman gain that assimilates the edge related measurements. It has to be noted that only one iteration per volume is performed to limit the computational complexity. Several edge detectors are supported by the framework, however in our case the step edge detector gave the most reliable results.

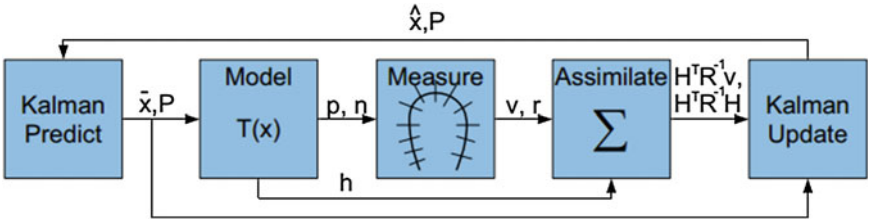


Fig. 2. Kalman filter processing chain including the prediction and update stages. [7]

The quality of the model fit is also computed by analyzing the number of edge outliers (locations where the edge detection failed) divided by the total number of edges:

$$N_{fit} = 1.0 - \frac{N_{outliers}}{N_{edges}} \tag{3}$$

Three anatomic landmarks are defined in the model's (u, v) parametric space, these are the left ventricle's apex, base and the middle of the outflow tract as seen in Fig. 3.b. If a fit of good quality is available the position of the anatomic landmarks relative to the ultrasound probe is computed and denoted LV_A, LV_B and LV_{OT} respectively.

2.2 Model Deformation

A geometric model representing the outer layer of the heart has been created to match the end-diastolic phase. As visible in Fig. 3.a this is the most common depiction of the heart in anatomy textbooks and therefore familiar to medical students. The positions of the three anatomic landmarks computed by the tracking algorithm have been identified manually on the generic model and are labeled as H_A, H_B and H_{OT}. The vectors defining the long axis of the tracked and generic heart model are be defined as: $\overrightarrow{M}_{long\ axis} = LV_B - LV_A$ and $\overrightarrow{H}_{long\ axis} = H_B - H_A$.

In order to add the virtual heart model to the scene, it needs to be spatially aligned and temporally deformed according to the data present in the ultrasound volumes. The global 3D pose transformation matrix aligning the heart model to the tracked model is given by:

$$MDM = T_{apex} * S_{long\ axis} * R_{outflow} * R_{long\ axis} \tag{4}$$

with $R_{long\ axis}$ the rotation matrix overlaying the vector $\overrightarrow{H}_{long\ axis}$ over $\overrightarrow{M}_{long\ axis}$ computed with the Rodrigues formula:

$$R_{long\ axis} = I + K * sin(\theta) + K^2 * (1 - cos(\theta)) \tag{5}$$

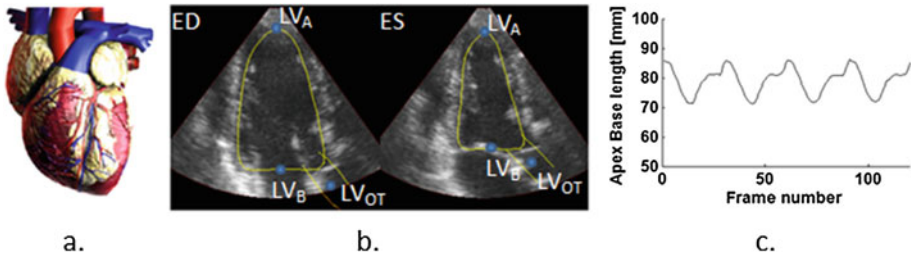


Fig. 3. (a.) Anatomic heart model. (b.) Fitted deformable model (yellow contour) at end diastole(ED) and end systole(ES), the position of the three anatomic landmarks is also indicated. (c.) Magnitude of $M_{long\ axis}$ over several heart cycles for a healthy volunteer, which is used for deforming the heart model (Color figure online).

with I the identity matrix, θ the angle between $\overrightarrow{M}_{long\ axis}$ and $\overrightarrow{H}_{longaxis}$ and K of the form:

$$K = \begin{bmatrix} 0 & -u_z & u_y & 0 \\ u_z & 0 & -u_x & 0 \\ -u_y & u_x & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{6}$$

where $\overrightarrow{u}(x, y, z)$ is the unit vector corresponding to $\overrightarrow{M}_{longaxis}$. Similarly $R_{outflow}$ is the rotation matrix that rotates the heart model around its long axis until H_{OT} and M_{OT} coincide. Finally, $S_{long\ axis}$ scales the heart mesh to fit with the tracked mesh (i.e. the length of $\overrightarrow{H}_{longaxis}$ matches the length of $\overrightarrow{M}_{longaxis}$), while T_{apex} is a translation to align H_A and LV_A.

As mentioned the heart model is static and represents only the end diastolic case, therefore a simplified kinematic model approximating the movement of the outer layer of the heart and matching the movement of the left ventricle extracted from the ultrasound data is implemented. Assuming that the apex of the heart remains stationary and the volume of the heart is nearly constant, the local deformation pattern of a vertex with regards to its position at time t_0 can be approximated as follows:

$$V = V_0 - \Delta_{AB} \frac{(V_0 - H_A) \cdot \overrightarrow{M}_{longaxis}}{||\overrightarrow{M}_{longaxis}||^2} \overrightarrow{u} \tag{7}$$

with $\cdot$ the dot product and $||\overrightarrow{v}||$ the magnitude of the vector. Δ_{AB} is:

$$\Delta_{AB} = max(||\overrightarrow{M}_{long\ axis}||) - ||\overrightarrow{M}_{long\ axis}|| \tag{8}$$

A typical pattern for Δ_{AB} is shown in the Fig. 3.c and correlates well with the motion of the heart's mitral plane (AV plane). The local movement of each vertex is represented as a translation matrix and is denoted MLD for convenience.

2.3 Patient and Probe Visual Tracking

To generate a consistent augmented view the position of the patient and of the ultrasonic probe with regards to the real world image acquired with the tablet's camera are required. We opted to use the Vuforia framework for frame marker tracking purposes. The position of two fiduciary markers, one attached to the patient and one attached to the probe are located in the 3D space of the tablet's camera.

Vuforia assumes that the camera is located at the origin and looks down on the Z axis. The pose of each visible marker with regards to the camera is obtained from Vuforia as a 4×4 homogeneous matrix and denoted MVM (Model View Matrix). To obtain the position of the a marker in the camera image the perspective projection matrix is needed, this is computed as:

$$PPM = \begin{bmatrix} tan^{-1}(\frac{FOV_x}{2}) & 0 & 0 & 0 \\ 0 & tan^{-1}(\frac{FOV_y}{2}) & 0 & 0 \\ 0 & 0 & -\frac{Z_{far}+Z_{near}}{Z_{far}-Z_{near}} & -\frac{2*Z_{far}*Z_{near}}{Z_{far}-Z_{near}} \\ 0 & 0 & -1 & 0 \end{bmatrix} \quad (9)$$

with FOV_x and FOV_y the field of view of the tablet's camera along x and y computed by Vuforia based on the camera calibration matrix and Z_{near}, Z_{far}, which represent the position of the near and far clipping planes along the Z axis.

Vuforia is also able to generate an augmented view by adding a set of geometric objects to the camera image, however we chose to retrieve the camera image stream from Vuforia and implement a custom augmentation scheme as described below.

2.4 Augmented View Generation

A customized version of VES/Kiwi, the embedded visualization toolbox from Kitware Inc., has been implemented for augmentation purposes. VES is a set of libraries that conveniently links VTK and OpenGL ES 2.0. By default the standard VTK camera model is implemented, however we extended VES to add support for the Vuforia camera model. Each of the elements in the scene has its custom OpenGL ES vertex and fragment shaders. A set of uniform values can be passed along to each object.

The augmented virtual scene is composed of the camera image, an ultrasound image plane, a triangle mesh with custom vertex deformation representing the high level anatomic heart model and a set of triangle meshes denoting anatomic structures of the heart (in our case the left ventricle and the simplified model of the outflow tract).

The position of each vertex $V(x, y, z, 1)$ in the augmented scene is given by:

$$gl\_Position = VAR * PPM * MVM * MCM * V; \quad (10)$$

with PPM and MVM as described before, MCM being a model calibration matrix giving the relative position of a scene element with regards to the origin of

its corresponding frame marker. And VAR, the view aspect ratio matrix which gives the relationship between the camera and the view-port of the device.

The heart model deformation described in Sect. 2.2 has been also implemented in the vertex shader. Since each vertex on the heart model undergoes a different deformation for each time frame this is a convenient approach. Assuming the position of a vertex at t_0 is V_0, it position at time t is:

$$V = MLD * MDM * V_0; \tag{11}$$

with MDM a model deformation matrix that aligns the heart model to the segmented deformable mesh and MLD a local deformation based on the position of the vertex along the apex to base axis.

The fragment shaders corresponding to scene elements are as follows. Texture lookup tables are used for the camera plane and the ultrasound image. Additionally, the camera image can be rendered semi-transparent or an elliptic region near the heart cut away to increase depth perception. The geometric meshes have a standard Gouraud shader attached, while the heart mesh has a normal map texture in addition to the color texture. As such fine geometric details are preserved despite the low resolution of the mesh, which is computationally beneficial. Clipping of the mesh with the ultrasound plane is also supported.

2.5 Implementation Details

The Kalman filter is implemented as a custom plugin on the ultrasound scanner and the resulting segmented mesh, anatomic landmarks, tracking score and the ultrasound image plane corresponding to the 0° elevation direction for each volume are sent to the tablet device via a TCP-IP socket.

Both an Android and an iOS application for augmentation purposes were created. They share the same native code and act as a wrapper on top of it to handle user input and the interface calls to Android and iOS respectively. Support for the tablet's front and back camera has been added. A Nexus 10 Samsung tablet has been used during the echo-lab tests.

Initially the heart frame marker is placed in a position as close as possible to the true position of the heart. However once 3D ultrasound data is available the MCM matrix for the heart model is updated to correctly place the heart with regards to the position of the frame maker. The MCM matrix for the probe containing the offset from the center of the frame maker to the center point on the probe's surface is computed empirically and is kept constant for all experiments. To avoid undesirable augmentation results given by poor input data, a threshold value of 20 % is applied to the tracking score N_{fit} in Eq. 3. If the tracking score for the current frame is below the threshold no augmentation is performed.

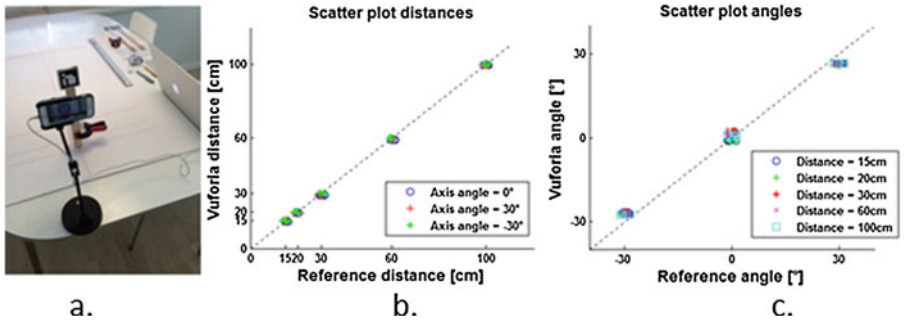


Fig. 4. (a.) Setup for evaluating the accuracy of Vuforia for visual marker tracking. Scatter plots for the distance along the Z axis (b.) and the angle along the Y axis (c.)

3 Experimental Results

3.1 Frame Maker Tracking Accuracy

The feasibility of Vuforia for frame marker tracking purposes was evaluated in an in-vitro study. Both the depth and rotational accuracy of the system were tested by placing a frame marker at 15, 20, 30, 60, 100 cm along the Z axis and at three different Y axis angles (0° and ±30°) under ambient lighting. The value ranges for depth and elevation angle were chosen since they correspond well with the typical ranges of our application. The setup used for evaluation as well as the angle and distance scatter plots are presented in the Fig. 4. For each distance and angle the test was repeated 5 times. The tracking accuracy is consistently good over the entire range of depth values with the error mean and standard deviation of $-0.31\,\text{cm} \pm 0.38\,\text{cm}$. The mean angle error and standard deviation are $-0.05° \pm 1.77°$, $2.91° \pm 0.29°$ and $-3.39° \pm 0.15°$ for an Y axis angle of 0°, −30° and 30° respectively. These values are reasonable when compared to the 2° elevation resolution of the cardiac probe.

3.2 In-vivo Laboratory Scanning

Figure 5.a illustrates a typical scanning scenario in the ultrasound lab. A Vivid E9 ultrasound scanner (GE Vingmed, Norway) has been used to acquire 3D apical views of the left ventricle. The tests were done without ECG triggering, during normal breathing or at end expiratory breath-hold.

Depending on the depth and the sector angle a volume rate between 6 and 12 frames per second was achievable. Scanning was performed on three healthy volunteers, by an experienced echocardiographer. Deformable model fitting was successful for all cases and model fitting times were 5 ms per 3D volume. The update frame rate for VES was set at 60 frames/second. The system functioned in real-time and no significant latency between the ultrasound data shown on the scanner and the image present on the tablet was observed.

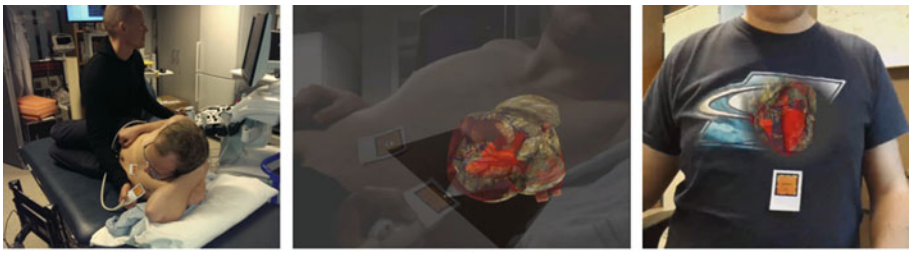


Fig. 5. (a.) Echo-lab setup for scanning a healthy volunteer. (b.) Screen capture from the tablet device illustrating the augmented view. (c.) Alternative augmentation that give a more accurate depth perception.

A sample augmented view, captured on the tablet device is presented in Fig. 5.b, whereas an improved augmented view of the heart model is shown is Fig. 5.c.

4 Discussion and Conclusion

The work presented an augmented reality framework that fuses an existing anatomical model of the heart with ultrasound images acquired in real time. The system was presented to an experienced cardiologist, and was deemed helpful for inexperienced users and students learning echocardiography. In contrast to other approaches it requires inexpensive hardware and it is portable since it is implemented on a tablet device. A highly informative view can be presented to the user allowing for easier correlation between cardiac anatomy and the acquired ultrasound image. By implementing the mesh deformation and augmentation on the tablet's GPU, a much faster response time to user interactions is attainable. Furthermore, the movement of the patient, probe and tablet's camera are continuously tracked and the augmented view is updated accordingly to maintain a consistent visualization, thus eliminating the need of manual alignment.

Vuforia was deemed accurate enough for tracking purposes, however the validation of the visual tracking accuracy was only an in-vitro study. Testing Vuforia's accuracy against an optical or magnetic tracker is planned.

Additionally, the system can receive a stream of data from a position sensor, which is, for example, attached to the probe. Integrating data from the position sensor was not used for this study. However it is of interest. Using such a sensor is beneficial to correct for rotation errors which increase once the angle between the frame maker and the camera is bigger than 30° or when the optical markers are not visible to the tablet's camera. Another problem that became apparent during scanning is that sometimes the probe does not face the tablet's camera. One solution would be to use marker-less tracking and instead cover the entire surface of the probe with a non repeating pattern. Vuforia has support for this

and early trials are underway. This would allow for 360° probe tracking and more flexibility during the acquisition.

Further extensions to the current prototype are envisioned. These include new applications such as fetal scanning or teaching tools for minimally invasive surgery. However, it is equally interesting to extend the prototype to work with 2D ultrasound images, the standard images obtained with portable scanners.

References

1. Bajura, M., Fuchs, H., Ohbuchi, R.: Merging virtual objects with the real world: seeing ultrasound imagery within the patient. ACM SIGGRAPH Comput. Graph. **26**(2), 203–210 (1992)
2. Bluteau, J., Kitahara, I., Kameda, Y., Noma, H., Kogure, K., Ohta, Y.: Visual support for medical communication by using projector-based augmented reality and thermal markers. In: Proceedings of the International Conference on Augmented Tele-Existence, pp. 98–105 (2005)
3. Ford, S., Viola, I., Bruckner, S., Torp, H., Kiss, G.: Heartpad: real-time visual guidance for cardiac ultrasound. In: Proceedings of the Workshop at SIGGRAPH Asia, pp. 169–176 (2012). http://www.cg.tuwien.ac.at/research/publications/2012/Ford-2012-HRV/
4. Harders, M., Bianchi, G., Knoerlein, B.: Multimodal augmented reality in medicine. In: Stephanidis, C. (ed.) UAHCI 2007 (Part II). LNCS, vol. 4555, pp. 652–658. Springer, Heidelberg (2007)
5. Kalkofen, D., Reitinger, B., Risholm, P., Bornik, A.: Integrated medical workflow for augmented reality applications. In: Proceedings of the International Workshop on Augmented environments for Medical Imaging and Computer-aided Surgery (2006)
6. Kimura, B.J., DeMaria, A.N.: Technology insight: hand-carried ultrasound cardiac assessment-evolution, not revolution. Nat. Clin. Pract. Cardiovasc. Med. **2**(4), 217–223 (2005). quiz 224. http://dx.doi.org/10.1038/ncpcardio0154
7. Orderud, F., Torp, H., Rabben, S.: Automatic alignment of standard views in 3D echocardiograms using real-time tracking. In: Procedings of the SPIE, vol. 7265, p. 72650D (2009)
8. State, A., Livingston, M., Garrett, W., Hirota., G., Whitton, M., Pisano, E., Fuchs, H.: Technologies for augmented reality systems: realizing ultrasound-guided needle biopsies. In: Proceedings of ACM SIGGRAPH, pp. 439–446 (1998)
9. Stetten, G., Chib, V.: Overlaying ultrasonographic images on direct vision. J. Ultrasound Med. **20**(3), 235–240 (2001)
10. Stetten, G., Chib, V., Tamburo, R.: Tomographic reflection to merge ultrasound images with direct vision. In: Proceedings of the IEEE Workshop on Applied Imagery Pattern Recognition, pp. 200–205 (2000)
11. Stüdeli, T., Kalkofen, D., Risholm, P., Ali, W., Freudenthal, A., Samset, E.: Visualization tool for improved accuracy in needle placement during percutaneous radiofrequency ablation of liver tumors. Proceedings of the SPIE, vol. 6918, p. 69180B (2008)

Author Index